US 20150382047A1

(19) **United States**

(12) **Patent Application Publication**

(10) **Pub. No.: US 2015/0382047 A1**

(43) **Pub. Date: Dec. 31, 2015**

VAN OS et al.

(54) **INTELLIGENT AUTOMATED ASSISTANT FOR TV USER INTERACTIONS**

(71) Applicant: **Apple Inc.**, Cupertino, CA (US)

(72) Inventors: **Marcel VAN OS**, San Francisco, CA (US); **Harry J. SADDLER**, Berkeley, CA (US); **Lia T. NAPOLITANO**, San Francisco, CA (US); **Jonathan H. RUSSELL**, Cupertino, CA (US); **Patrick M. LISTER**, Cupertino, CA (US); **Rohit DASARI**, Cupertino, CA (US)

(21) Appl. No.: **14/498,503**

(22) Filed: **Sep. 26, 2014**

Related U.S. Application Data

(60) Provisional application No. 62/019,312, filed on Jun. 30, 2014.

Publication Classification

(51) **Int. Cl.**

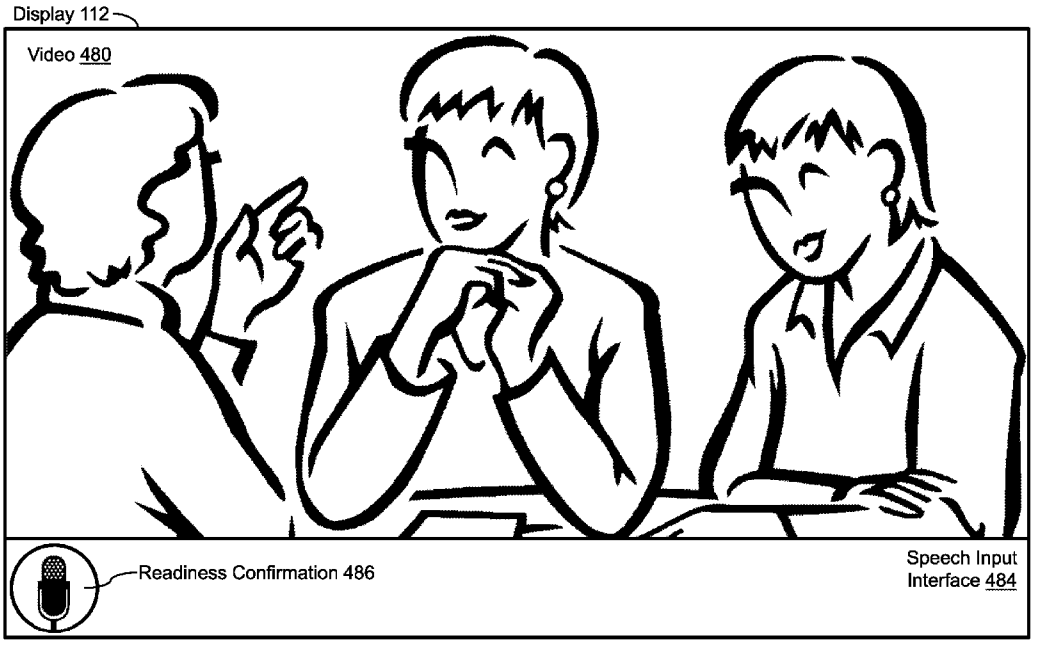
H04N 21/422	(2006.01)
H04N 21/443	(2006.01)
H04N 21/84	(2006.01)
H04N 21/858	(2006.01)
H04N 21/488	(2006.01)
H04N 21/41	(2006.01)
H04N 21/482	(2006.01)

(52) **U.S. Cl.**

CPC ***H04N 21/42203*** (2013.01); ***H04N 21/4126*** (2013.01); ***H04N 21/4438*** (2013.01); ***H04N 21/4828*** (2013.01); ***H04N 21/858*** (2013.01); ***H04N 21/4882*** (2013.01); ***H04N 21/84*** (2013.01)

(57) **ABSTRACT**

Systems and processes are disclosed for controlling television user interactions using a virtual assistant. A virtual assistant can interact with a television set-top box to control content shown on a television. Speech input for the virtual assistant can be received from a device with a microphone. User intent can be determined from the speech input, and the virtual assistant can execute tasks according to the user's intent, including causing playback of media on the television. Virtual assistant interactions can be shown on the television in interfaces that expand or contract to occupy a minimal amount of space while conveying desired information. Multiple devices associated with multiple displays can be used to determine user intent from speech input as well as to convey information to users. In some examples, virtual assistant query suggestions can be provided to the user based on media content shown on a display.



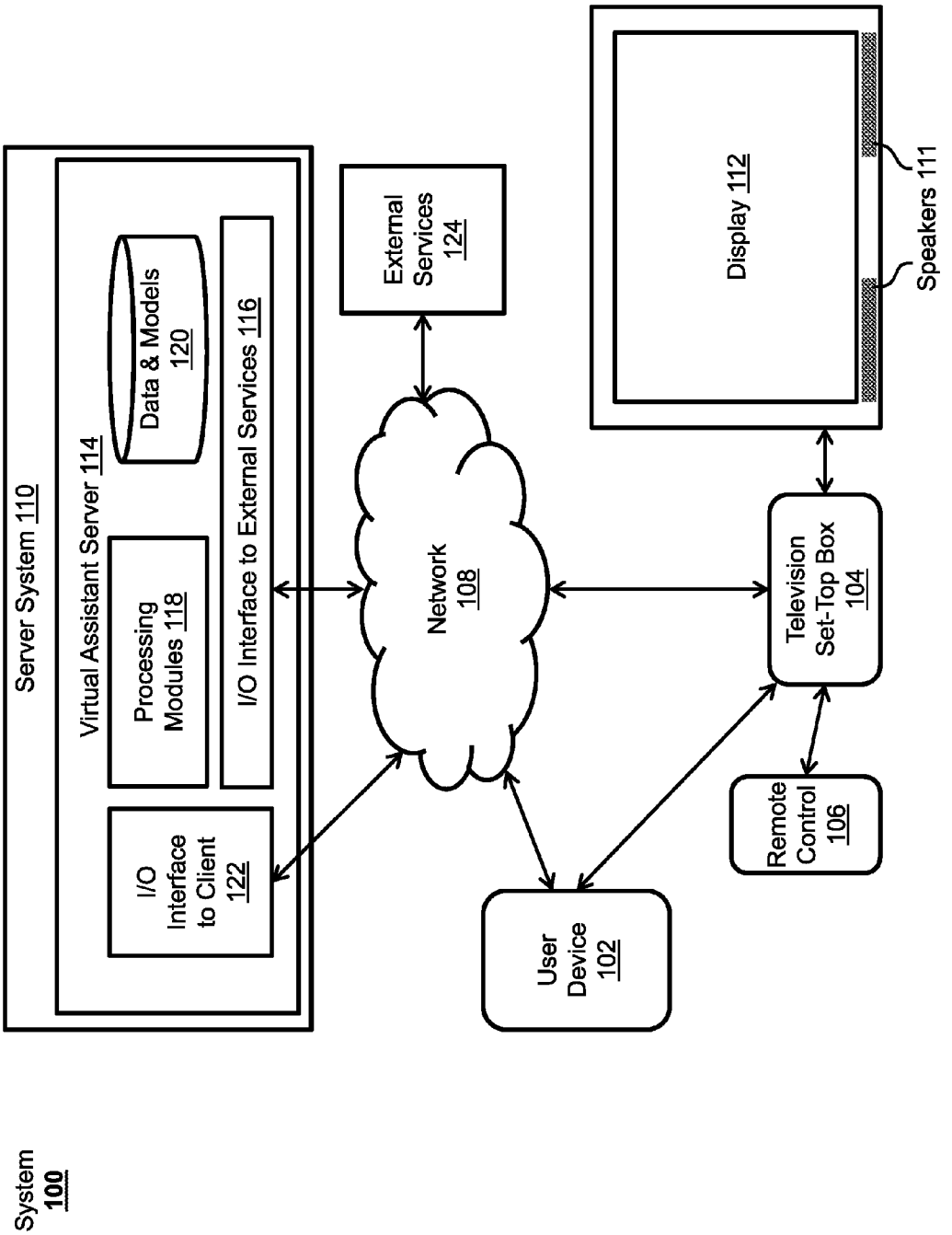

System
100
Server System 110
Virtual Assistant Server 114
Processing Modules 118
Data & Models 120
I/O Interface to Client 122
I/O Interface to External Services 116
External Services 124
Network 108
User Device 102
Television Set-Top Box 104
Remote Control 106
Display 112
Speakers 111

FIG. 1

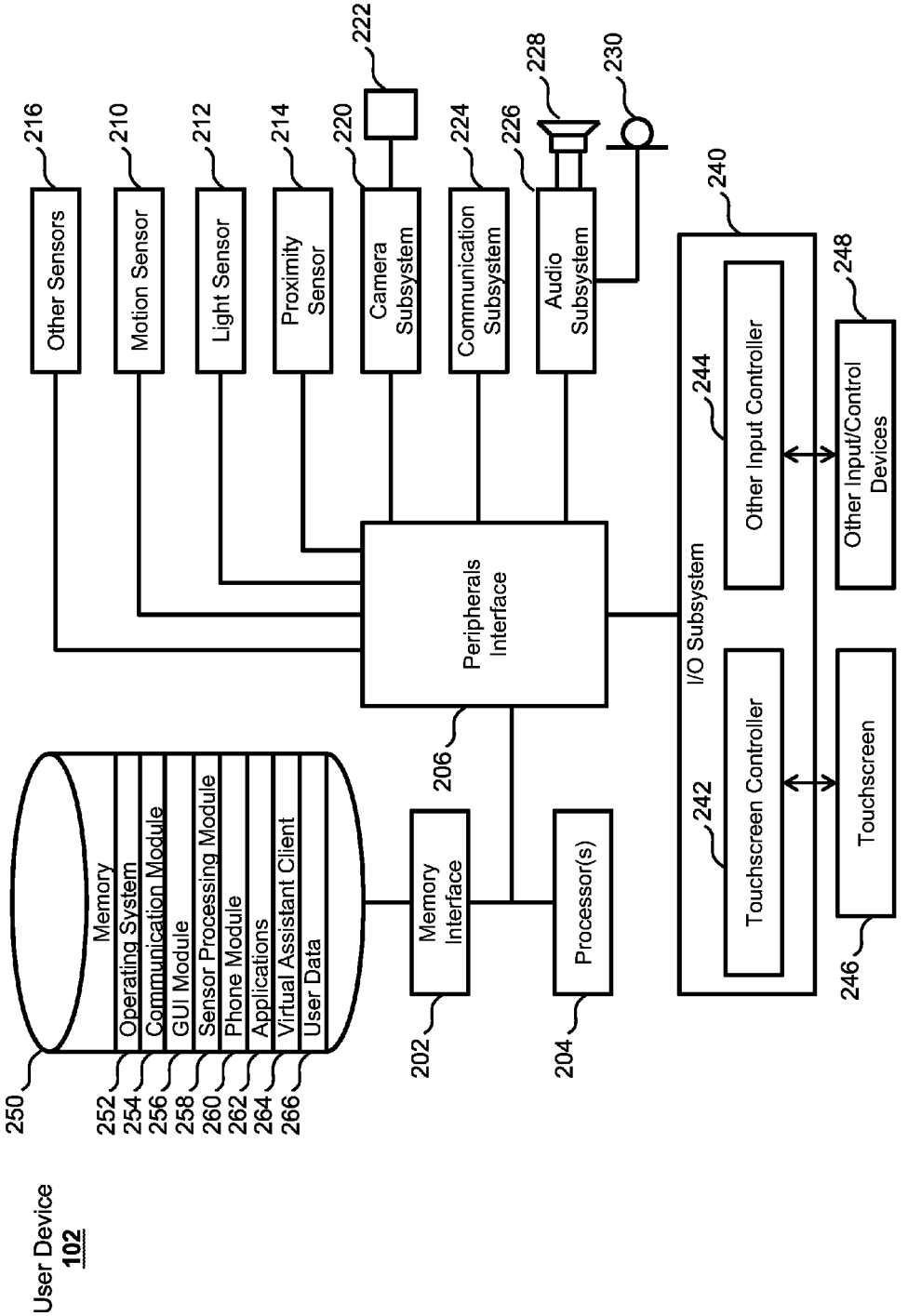

User Device
102
Other Sensors
216
Motion Sensor
210
Light Sensor
212
Proximity Sensor
214
Camera Subsystem
220
222
Communication Subsystem
224
Audio Subsystem
226
228
230
Peripherals Interface
206
I/O Subsystem
240
Other Input Controller
244
Other Input/Control Devices
248
Touchscreen Controller
242
Touchscreen
246
Memory
250
Operating System
252
Communication Module
254
GUI Module
256
Sensor Processing Module
258
Phone Module
260
Applications
262
Virtual Assistant Client
264
User Data
266
Memory Interface
202
Processor(s)
204

FIG. 2

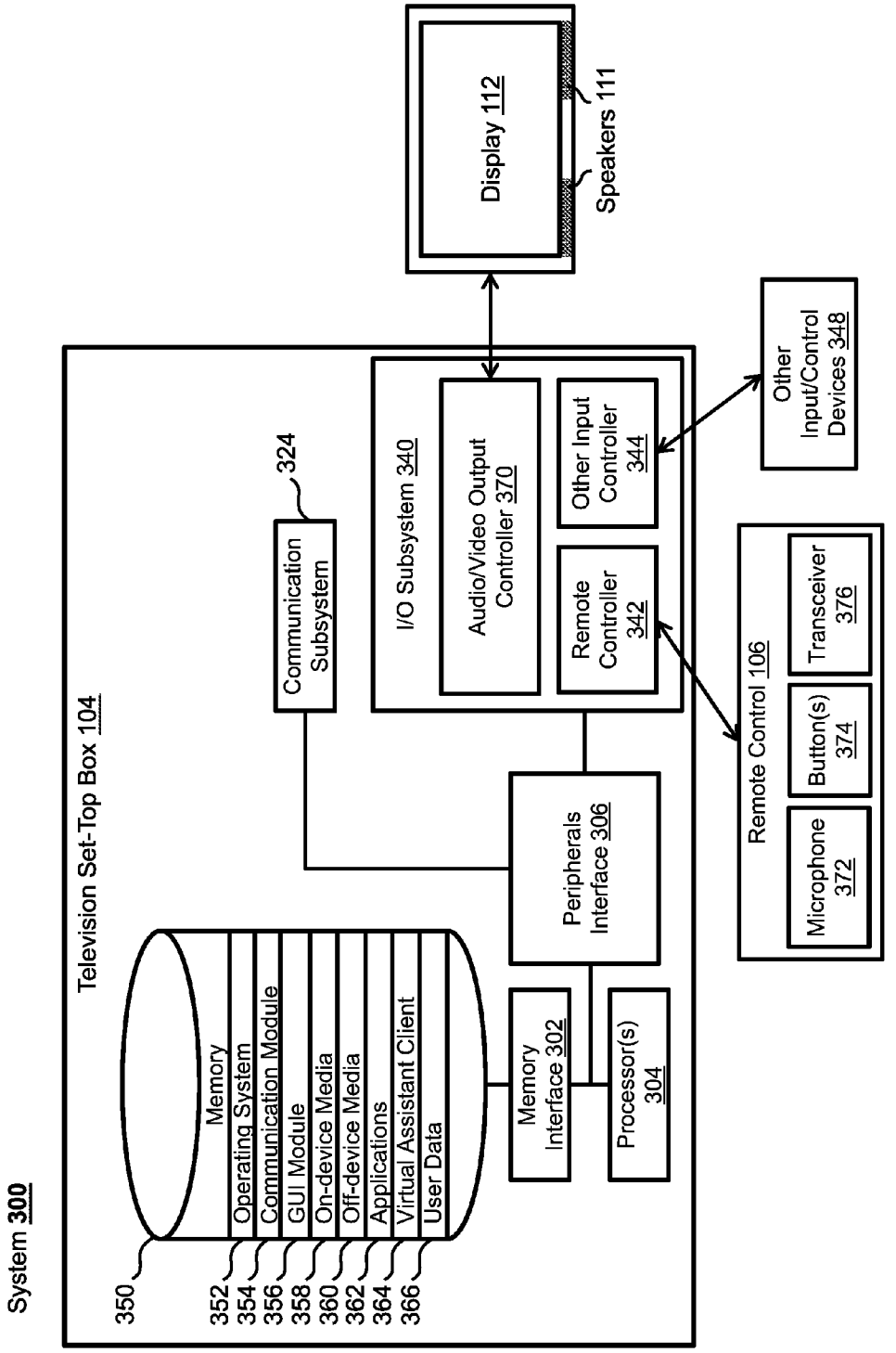

System 300
Television Set-Top Box 104
Memory
Operating System
Communication Module
GUI Module
On-device Media
Off-device Media
Applications
Virtual Assistant Client
User Data
350
352
354
356
358
360
362
364
366
Memory Interface 302
Processor(s) 304
Peripherals Interface 306
Communication Subsystem
324
I/O Subsystem 340
Audio/Video Output Controller 370
Other Input Controller 344
Remote Controller 342
Display 112
Speakers 111
Other Input/Control Devices 348
Remote Control 106
Microphone 372
Button(s) 374
Transceiver 376

FIG. 3

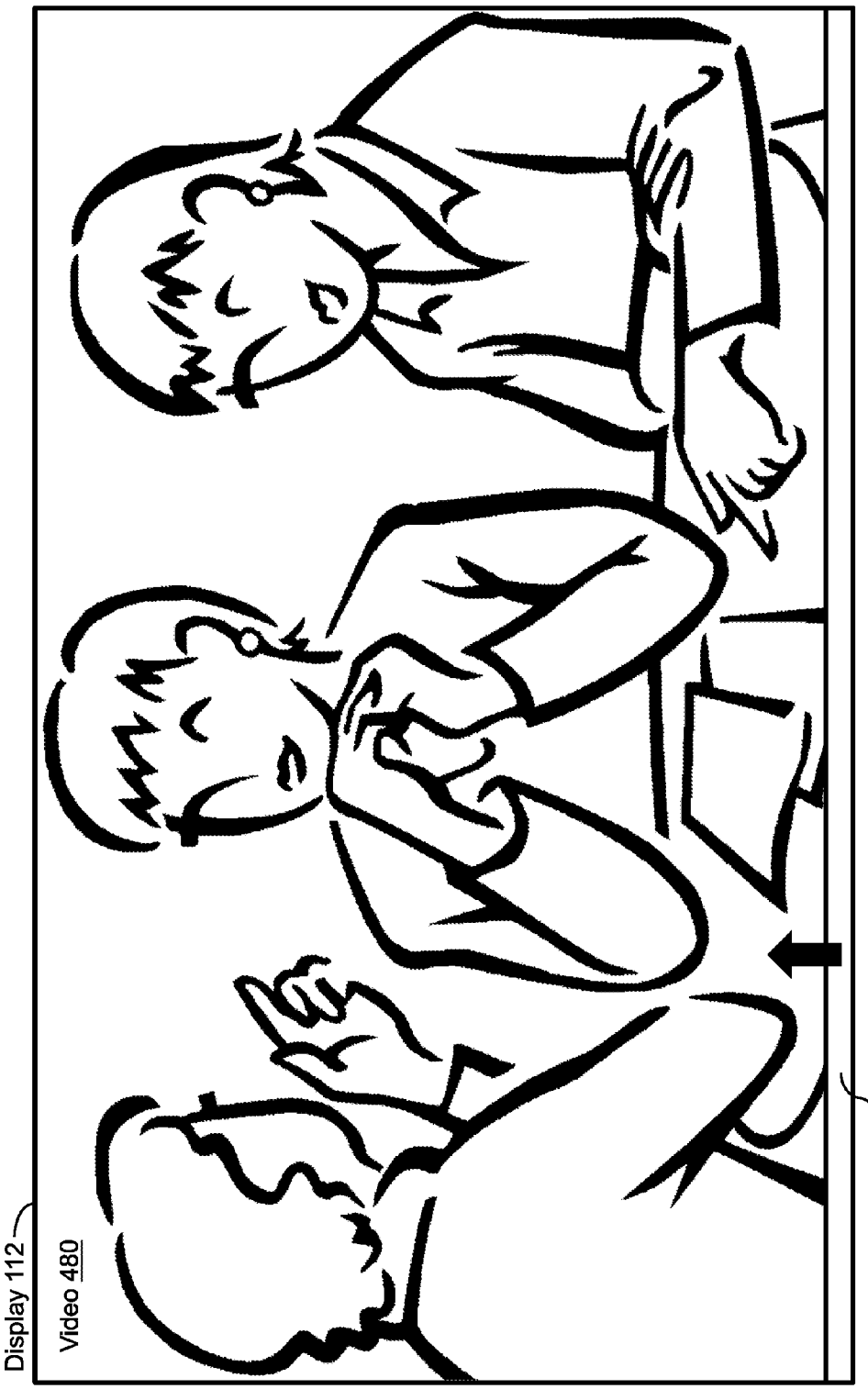


FIG. 4A

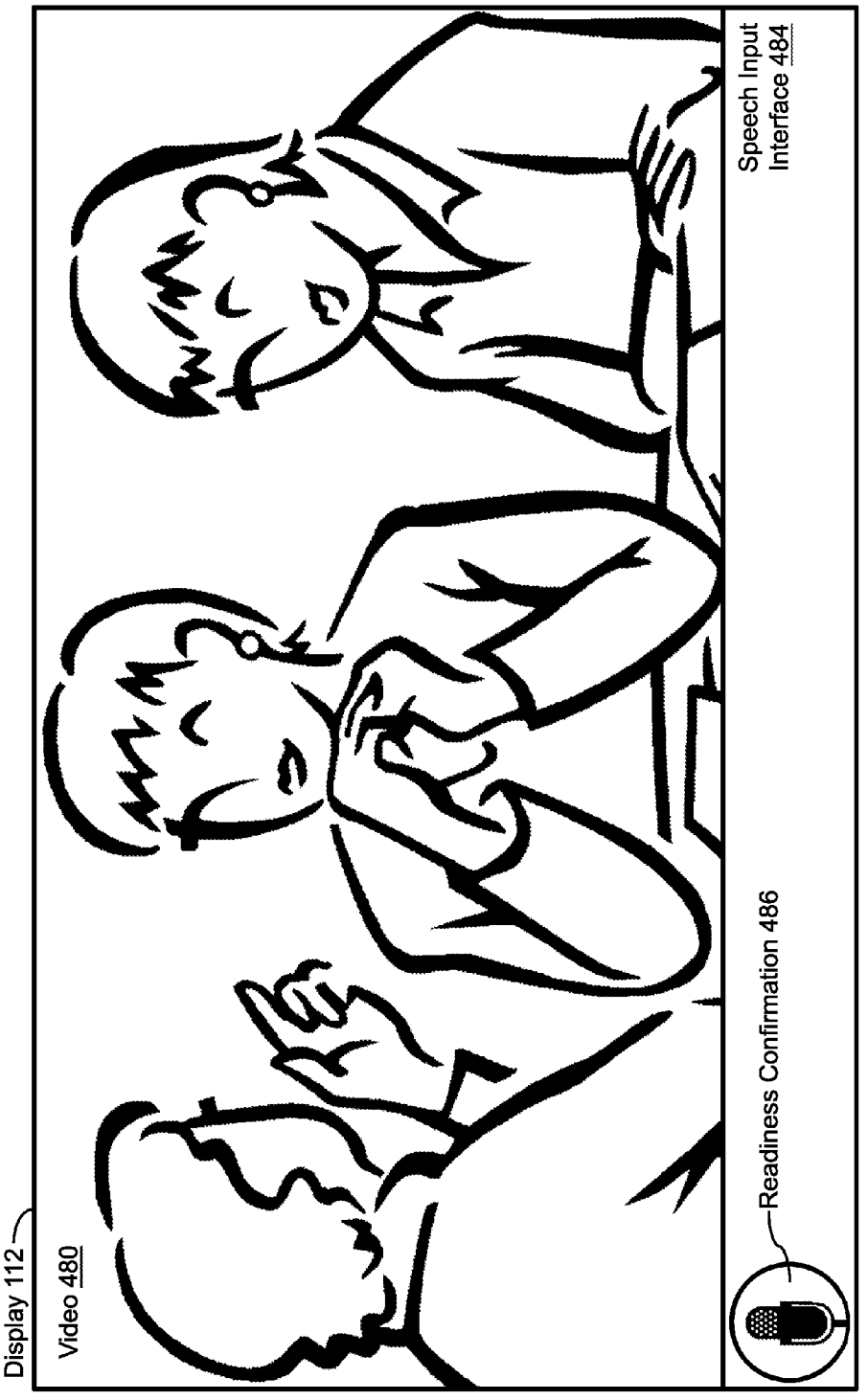

Display 112
Video 480
Speech Input Interface 484
Readiness Confirmation 486

FIG. 4B

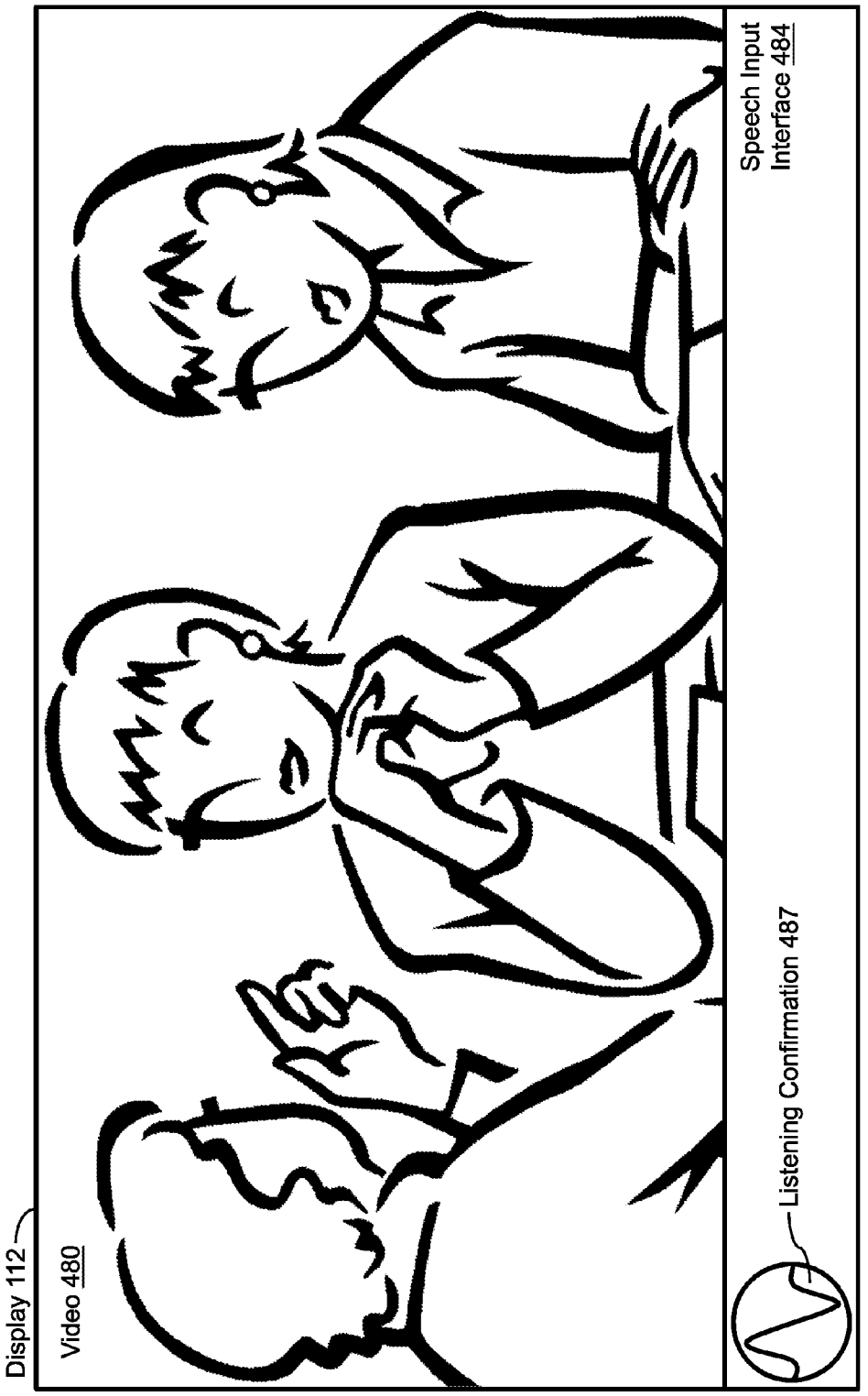

Display 112
Video 480
Speech Input Interface 484
Listening Confirmation 487

FIG. 4C

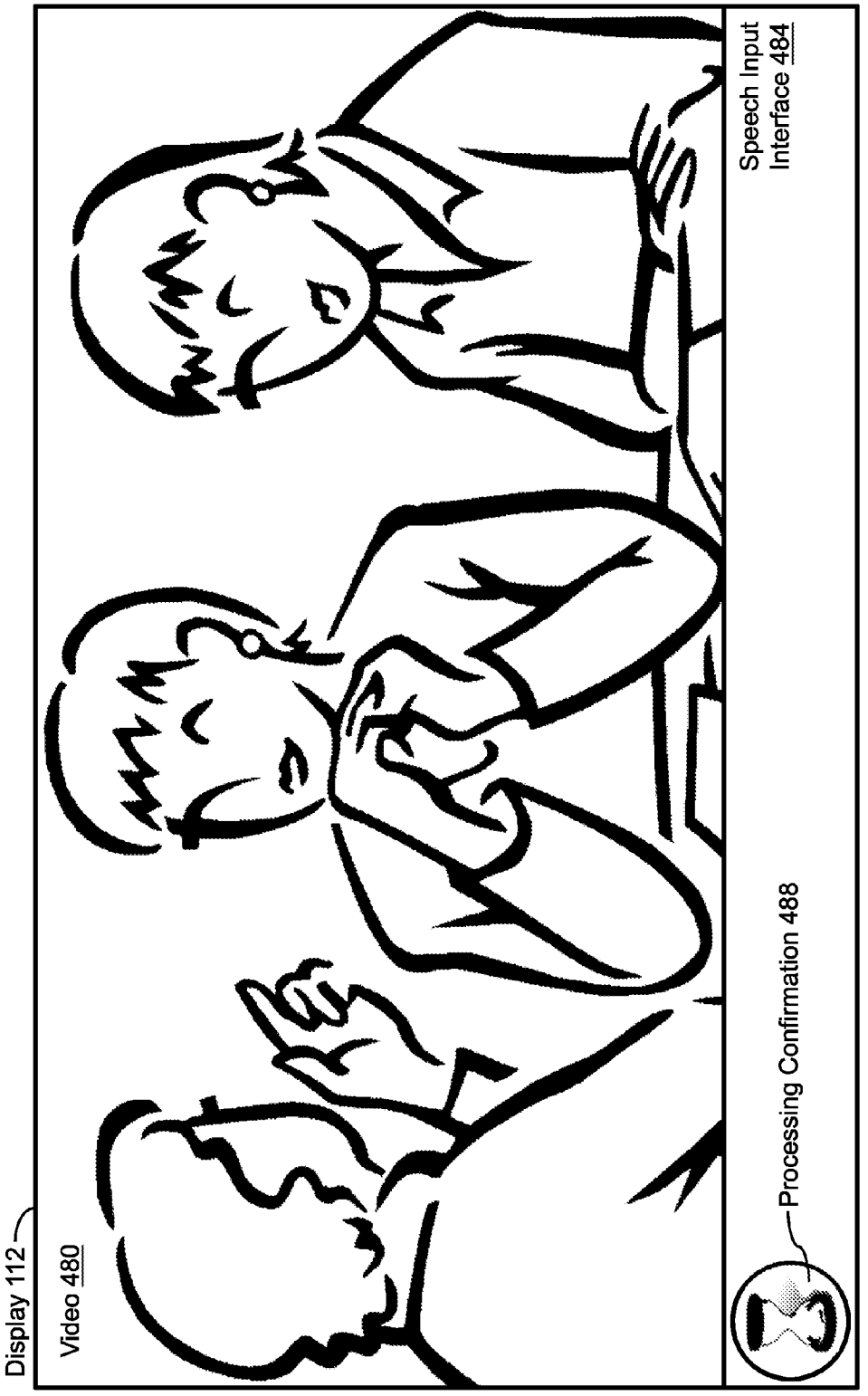

Display 112
Video 480
Speech Input Interface 484
Processing Confirmation 488

FIG. 4D

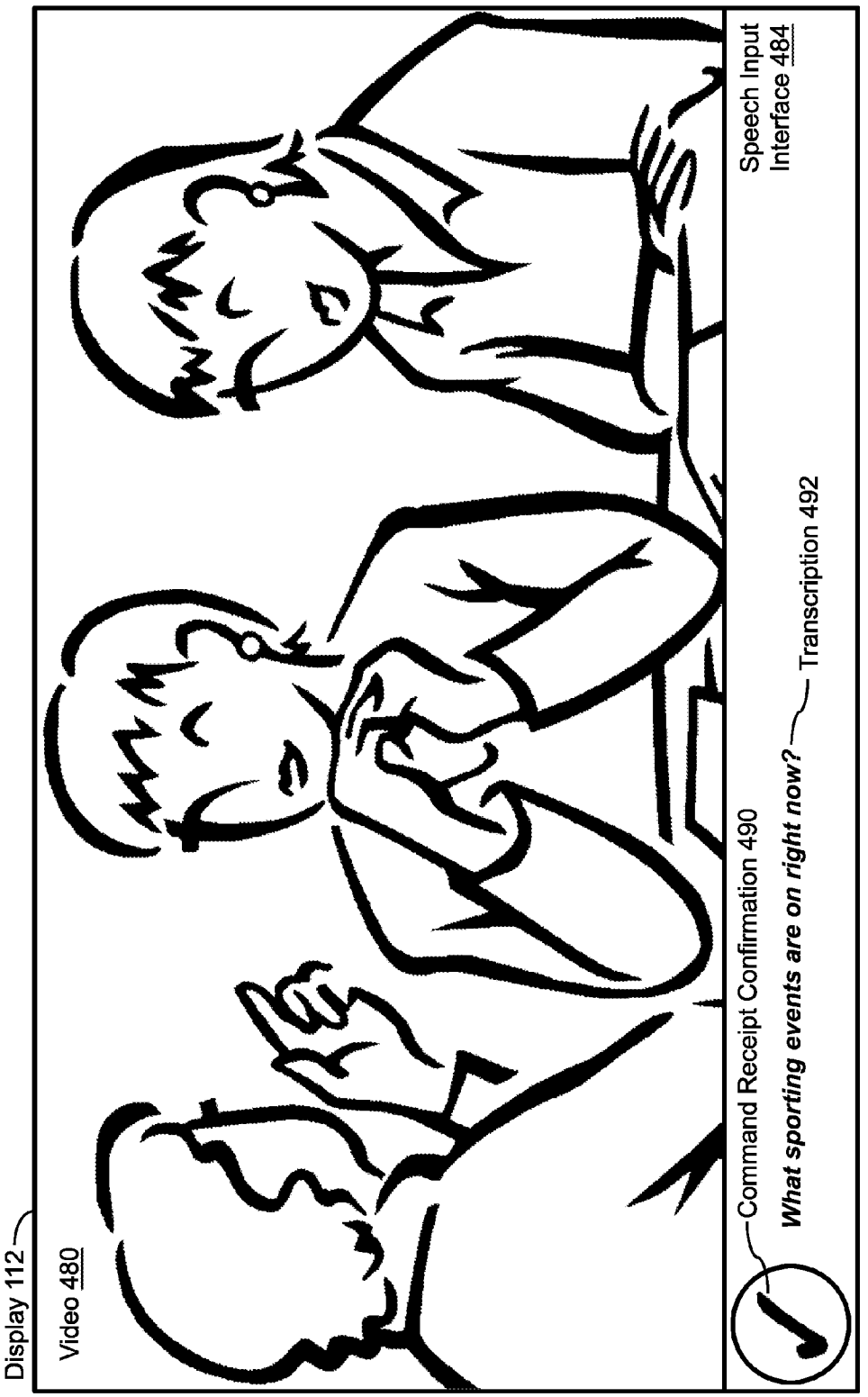

Display 112
Video 480
Speech Input Interface 484
Command Receipt Confirmation 490
What sporting events are on right now?
Transcription 492

FIG. 4E

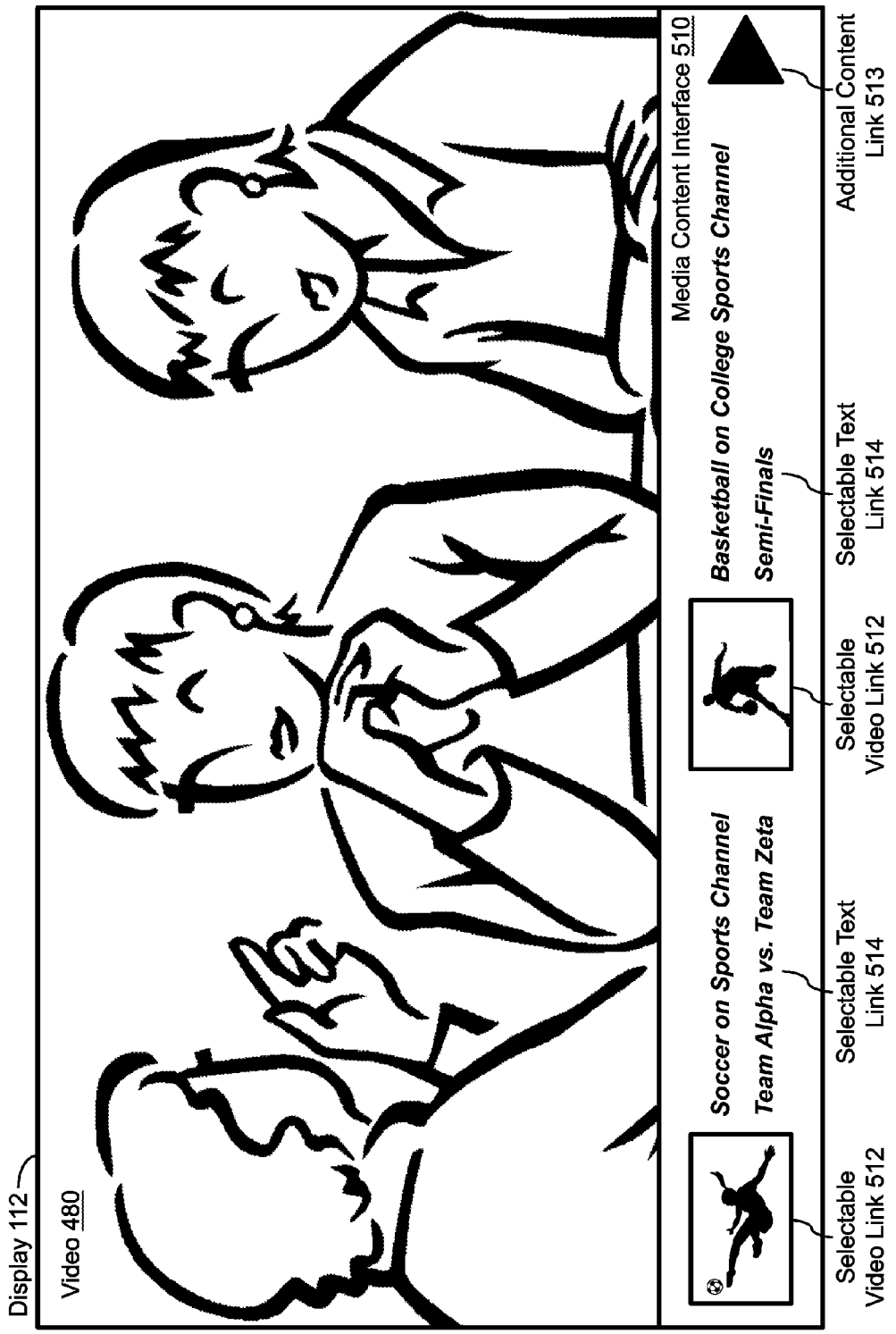

Display 112
Video 480
Media Content Interface 510
Soccer on Sports Channel
Team Alpha vs. Team Zeta
Basketball on College Sports Channel
Semi-Finals
Selectable Video Link 512
Selectable Text Link 514
Selectable Video Link 512
Selectable Text Link 514
Additional Content Link 513

FIG. 5

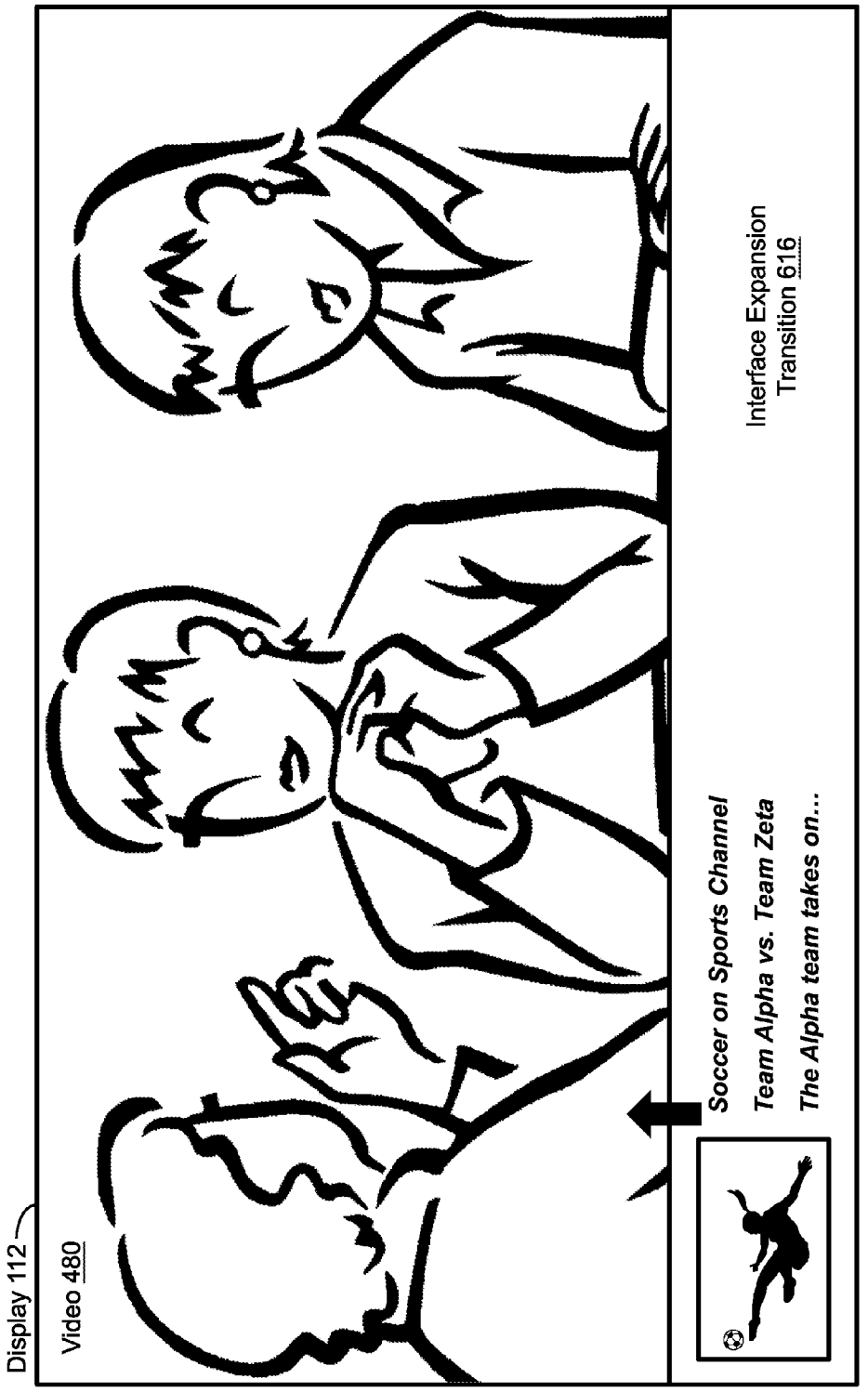

Display 112
Video 480
Soccer on Sports Channel
Team Alpha vs. Team Zeta
The Alpha team takes on...
Interface Expansion Transition 616

FIG. 6A

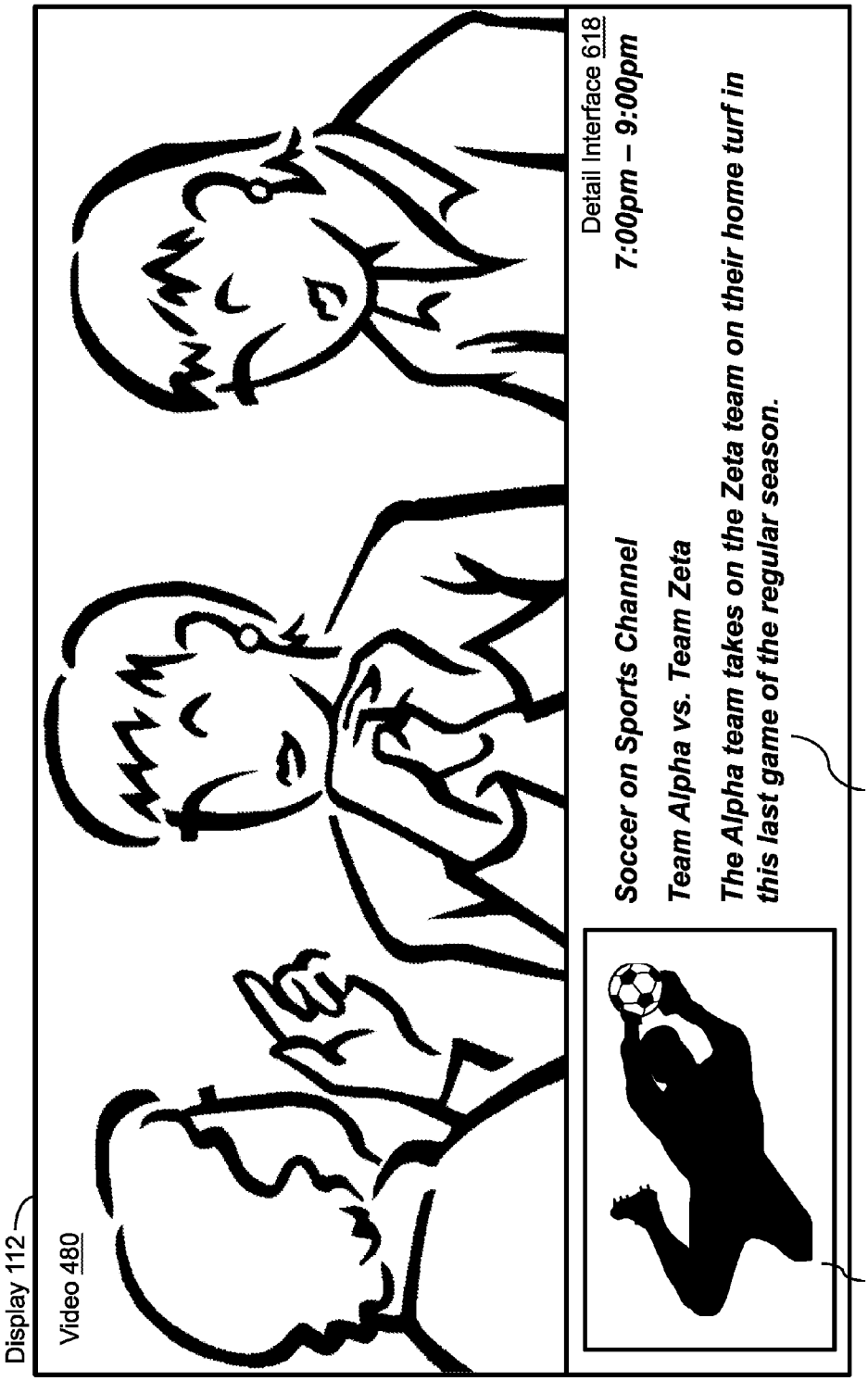

Display 112
Video 480
Detail Interface 618
7:00pm – 9:00pm
Soccer on Sports Channel
Team Alpha vs. Team Zeta
The Alpha team takes on the Zeta team on their home turf in this last game of the regular season.
Detailed Media Information 622
Selectable Video Link 620

FIG. 6B

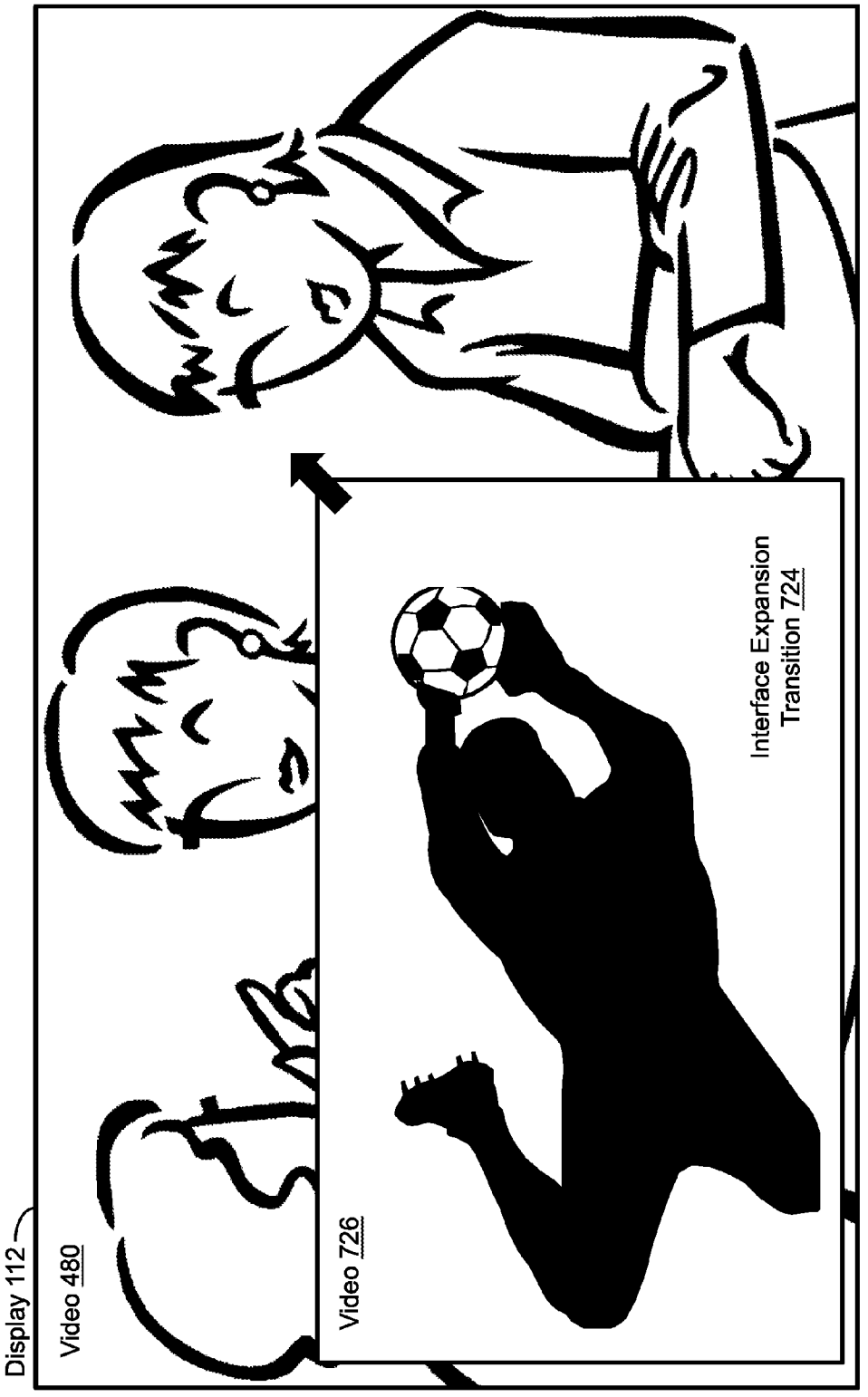

Display 112
Video 480
Video 726
Interface Expansion Transition 724

FIG. 7A

FIG. 7B

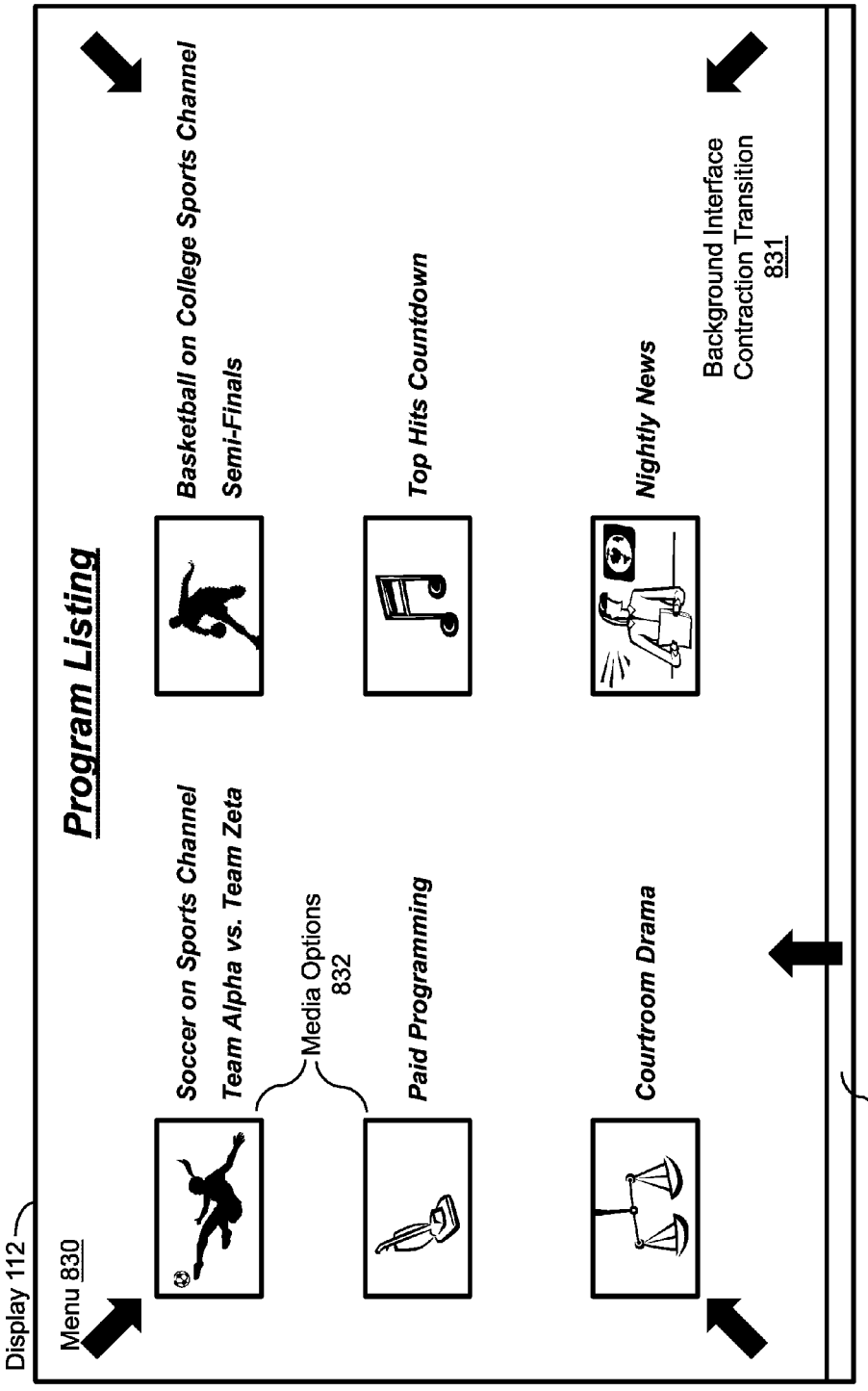

Display 112
Menu 830
Program Listing
Soccer on Sports Channel
Team Alpha vs. Team Zeta
Media Options
832
Paid Programming
Courtroom Drama
Basketball on College Sports Channel
Semi-Finals
Top Hits Countdown
Nightly News
Background Interface
Contraction Transition
831
Large Notification Area
834

FIG. 8A

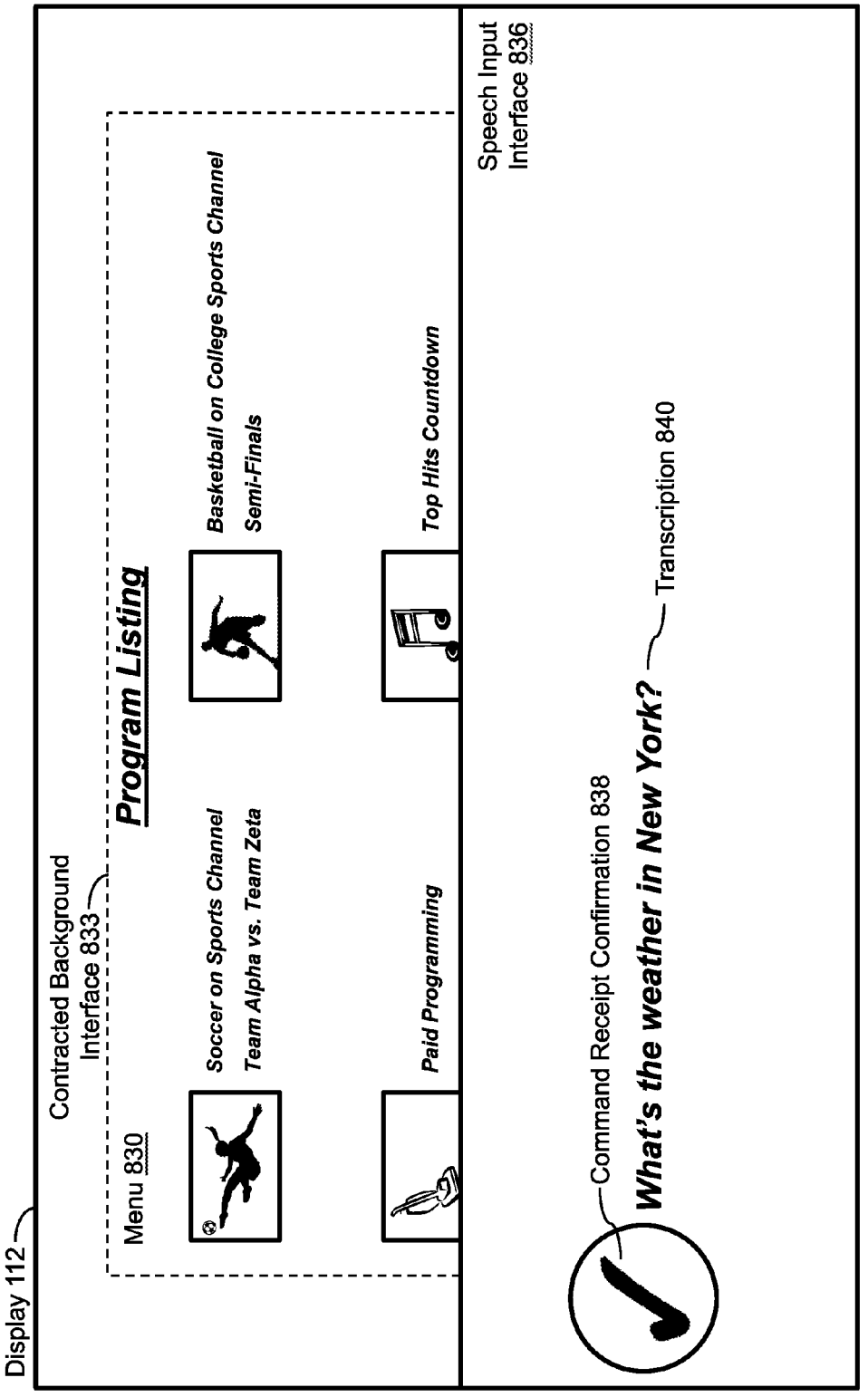

Display 112
Contracted Background Interface 833
Menu 830
Program Listing
Soccer on Sports Channel
Team Alpha vs. Team Zeta
Basketball on College Sports Channel
Semi-Finals
Paid Programming
Top Hits Countdown
Speech Input Interface 836
Command Receipt Confirmation 838
What's the weather in New York?
Transcription 840

FIG. 8B

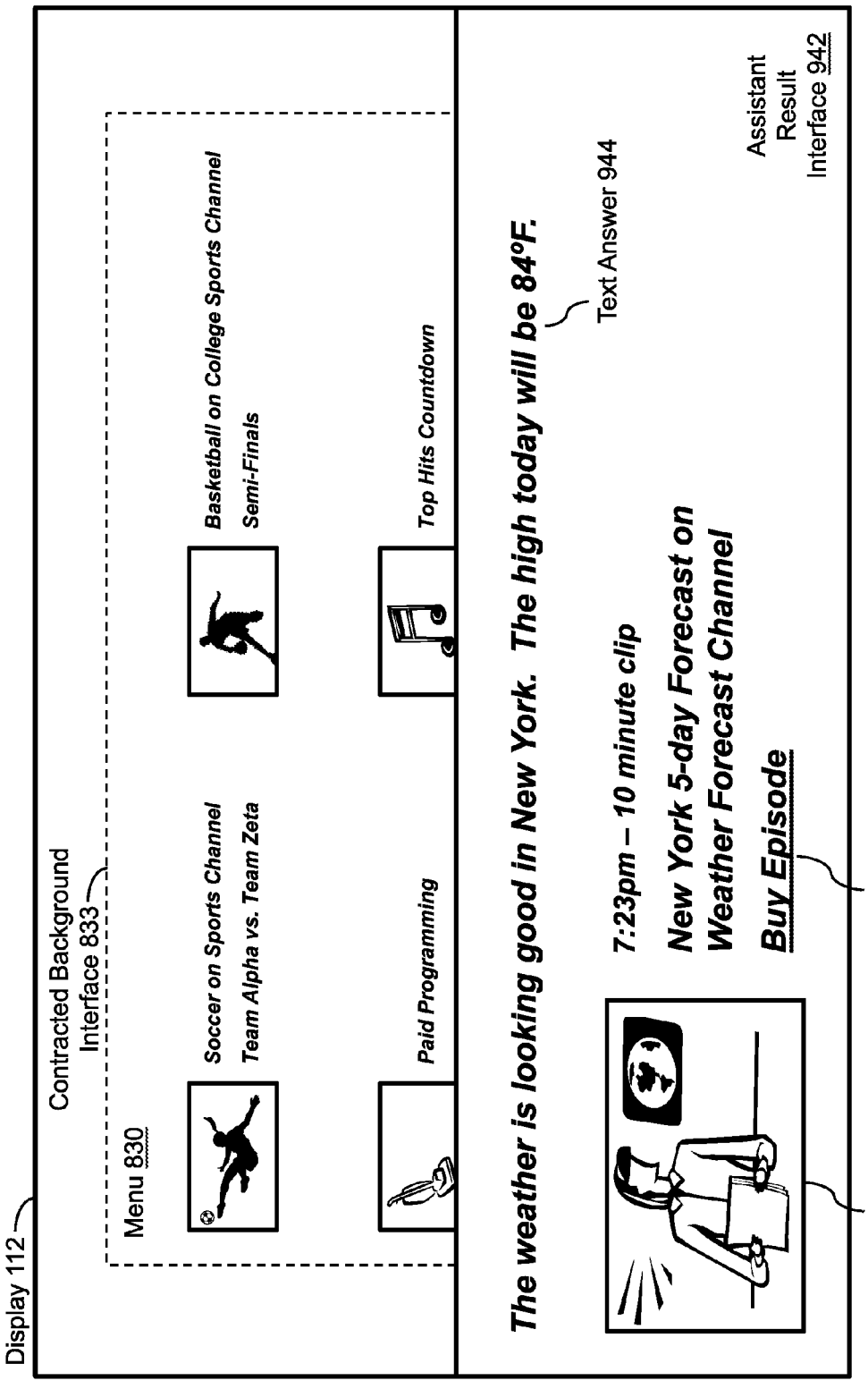

Display 112
Contracted Background Interface 833
Menu 830
Soccer on Sports Channel
Team Alpha vs. Team Zeta
Basketball on College Sports Channel
Semi-Finals
Paid Programming
Top Hits Countdown
The weather is looking good in New York. The high today will be 84°F.
Text Answer 944
7:23pm – 10 minute clip
New York 5-day Forecast on Weather Forecast Channel
Buy Episode
Assistant Result Interface 942
Selectable Video Link 946
Purchase Link 948

FIG. 9

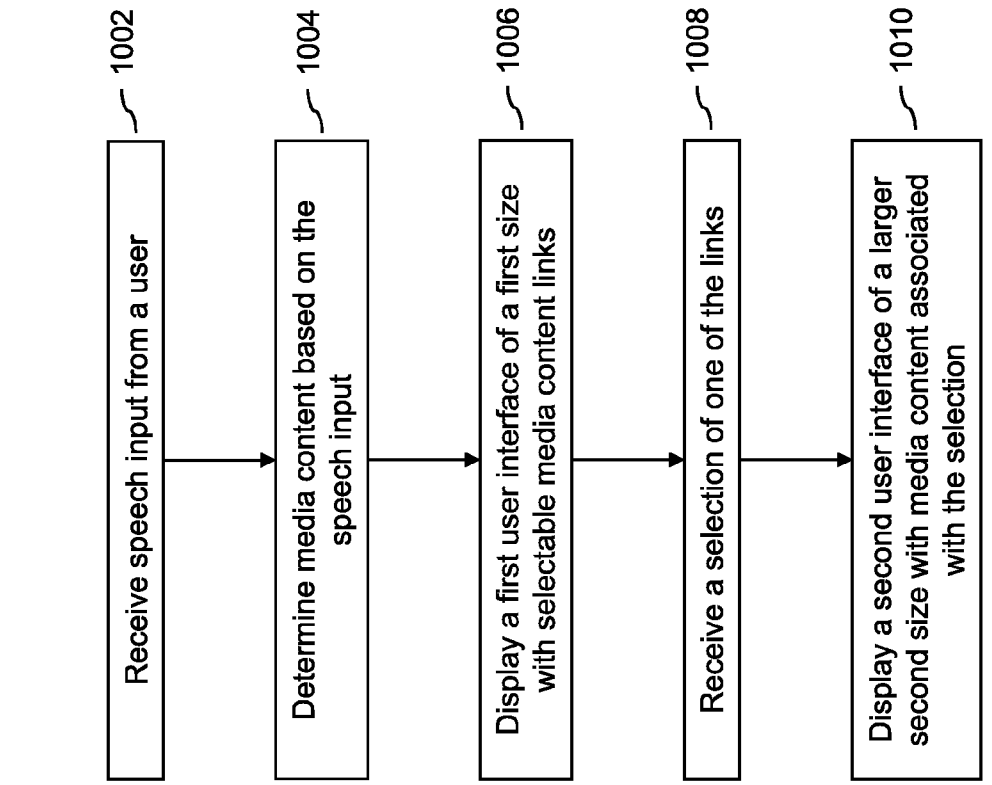

Process
1000
1002
Receive speech input from a user
1004
Determine media content based on the speech input
1006
Display a first user interface of a first size with selectable media content links
1008
Receive a selection of one of the links
1010
Display a second user interface of a larger second size with media content associated with the selection

FIG. 10

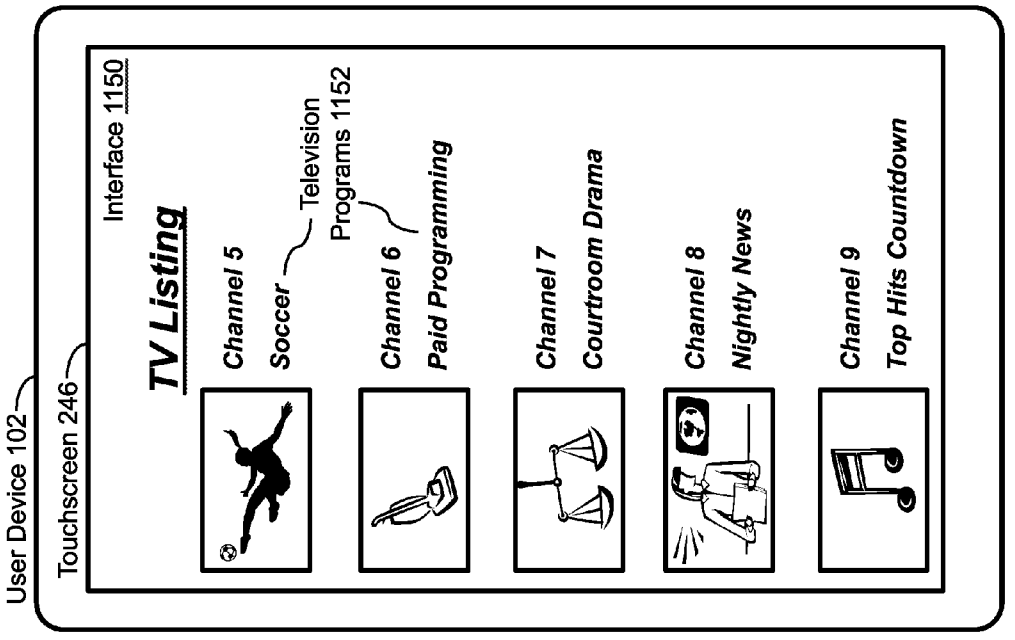

User Device 102
Touchscreen 246
Interface 1150
TV Listing
Channel 5
Soccer
Television Programs 1152
Channel 6
Paid Programming
Channel 7
Courtroom Drama
Channel 8
Nightly News
Channel 9
Top Hits Countdown

FIG. 11

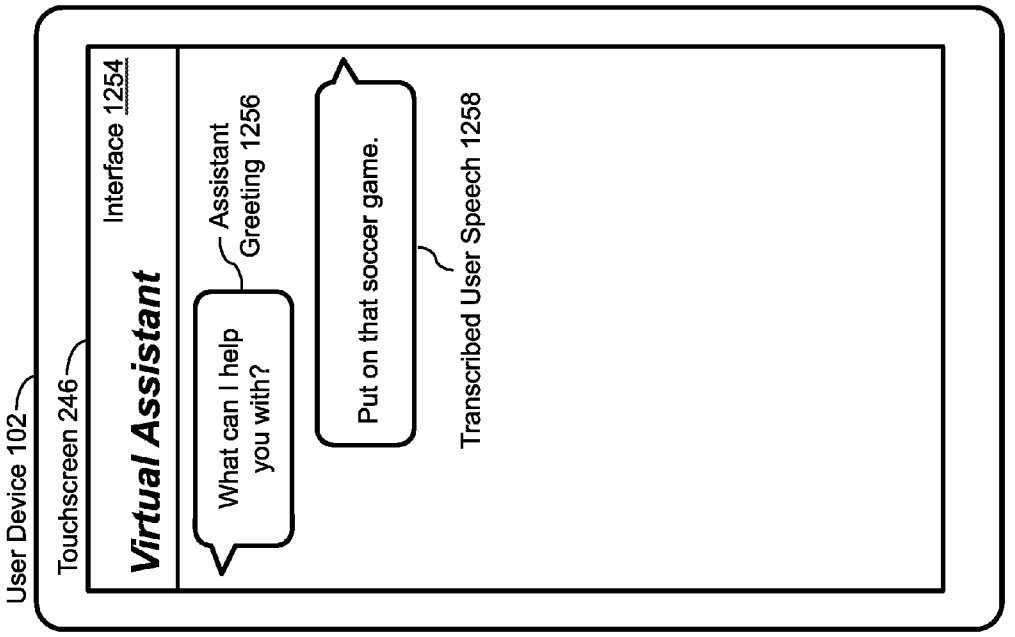

User Device 102
Touchscreen 246
Interface 1254
Virtual Assistant
What can I help you with?
Assistant Greeting 1256
Put on that soccer game.
Transcribed User Speech 1258

FIG. 12

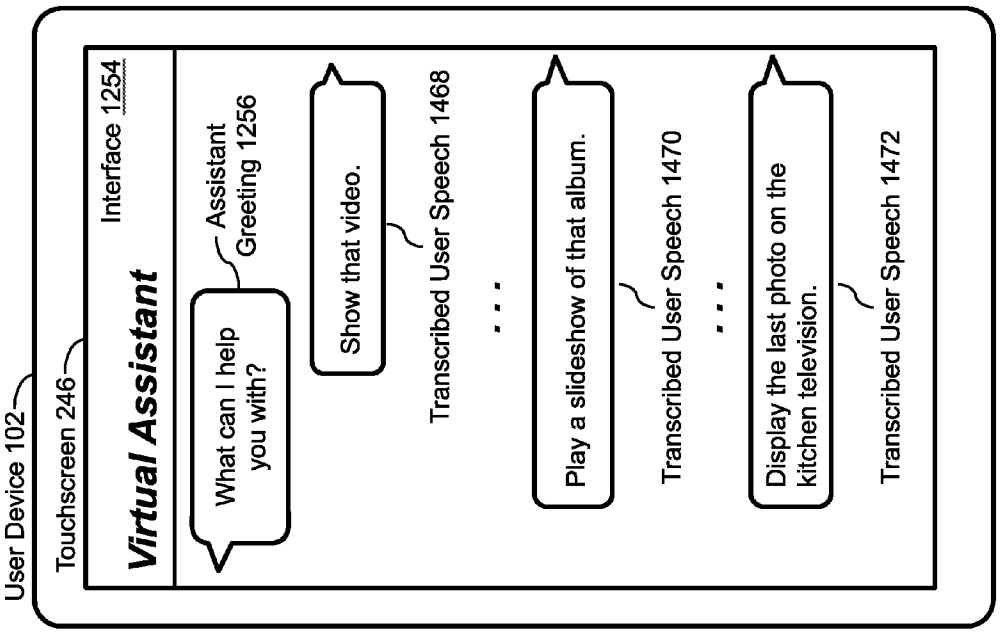

User Device 102
Touchscreen 246
Interface 1254
Virtual Assistant
Assistant Greeting 1256
What can I help you with?
Show that video.
Transcribed User Speech 1468
. . .
Play a slideshow of that album.
Transcribed User Speech 1470
. . .
Display the last photo on the kitchen television.
Transcribed User Speech 1472

FIG. 14

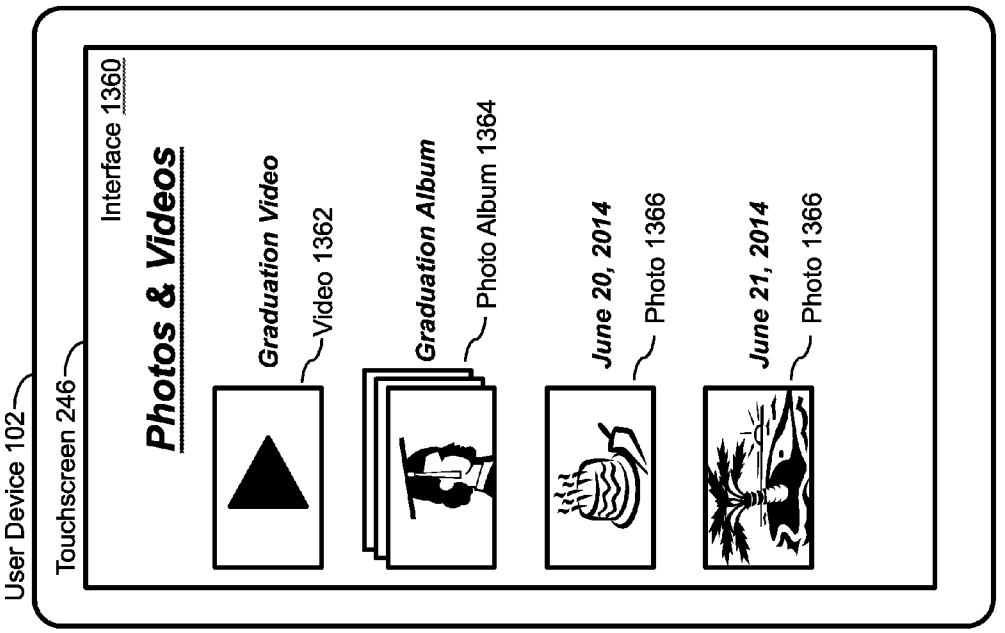

User Device 102
Touchscreen 246
Interface 1360
Photos & Videos
Graduation Video
Video 1362
Graduation Album
Photo Album 1364
June 20, 2014
Photo 1366
June 21, 2014
Photo 1366

FIG. 13

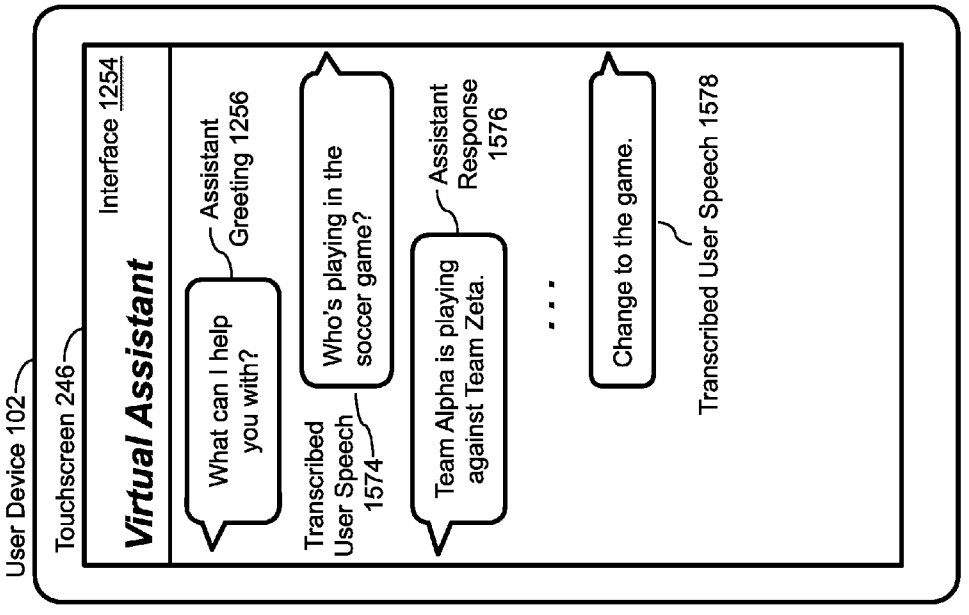

User Device 102
Touchscreen 246
Interface 1254
Virtual Assistant
What can I help you with?
Assistant Greeting 1256
Transcribed User Speech 1574
Who's playing in the soccer game?
Team Alpha is playing against Team Zeta.
Assistant Response 1576
. . .
Change to the game.
Transcribed User Speech 1578

FIG. 15

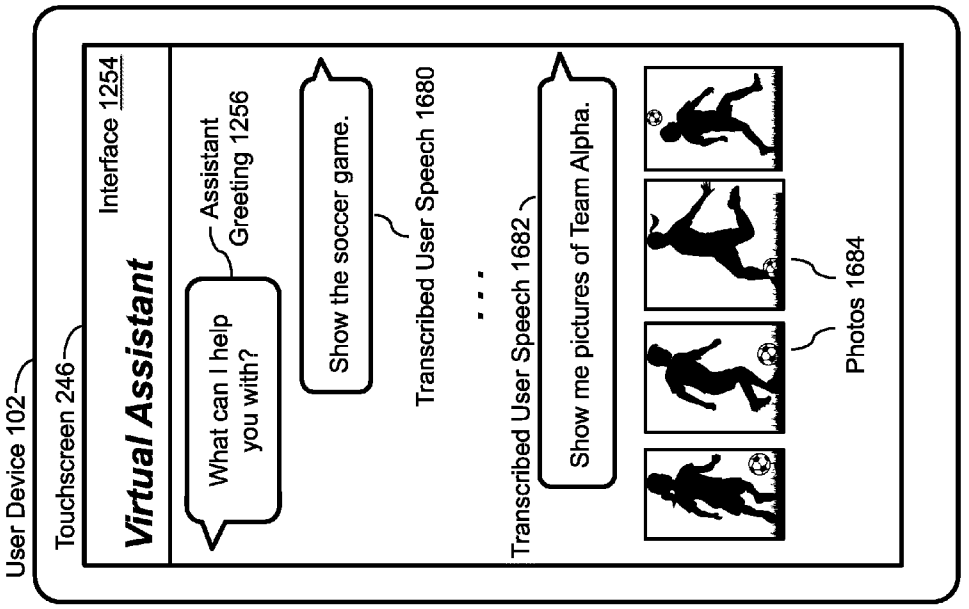

User Device 102
Touchscreen 246
Interface 1254
Virtual Assistant
What can I help you with?
Assistant Greeting 1256
Show the soccer game.
Transcribed User Speech 1680
. . .
Transcribed User Speech 1682
Show me pictures of Team Alpha.
Photos 1684

FIG. 16

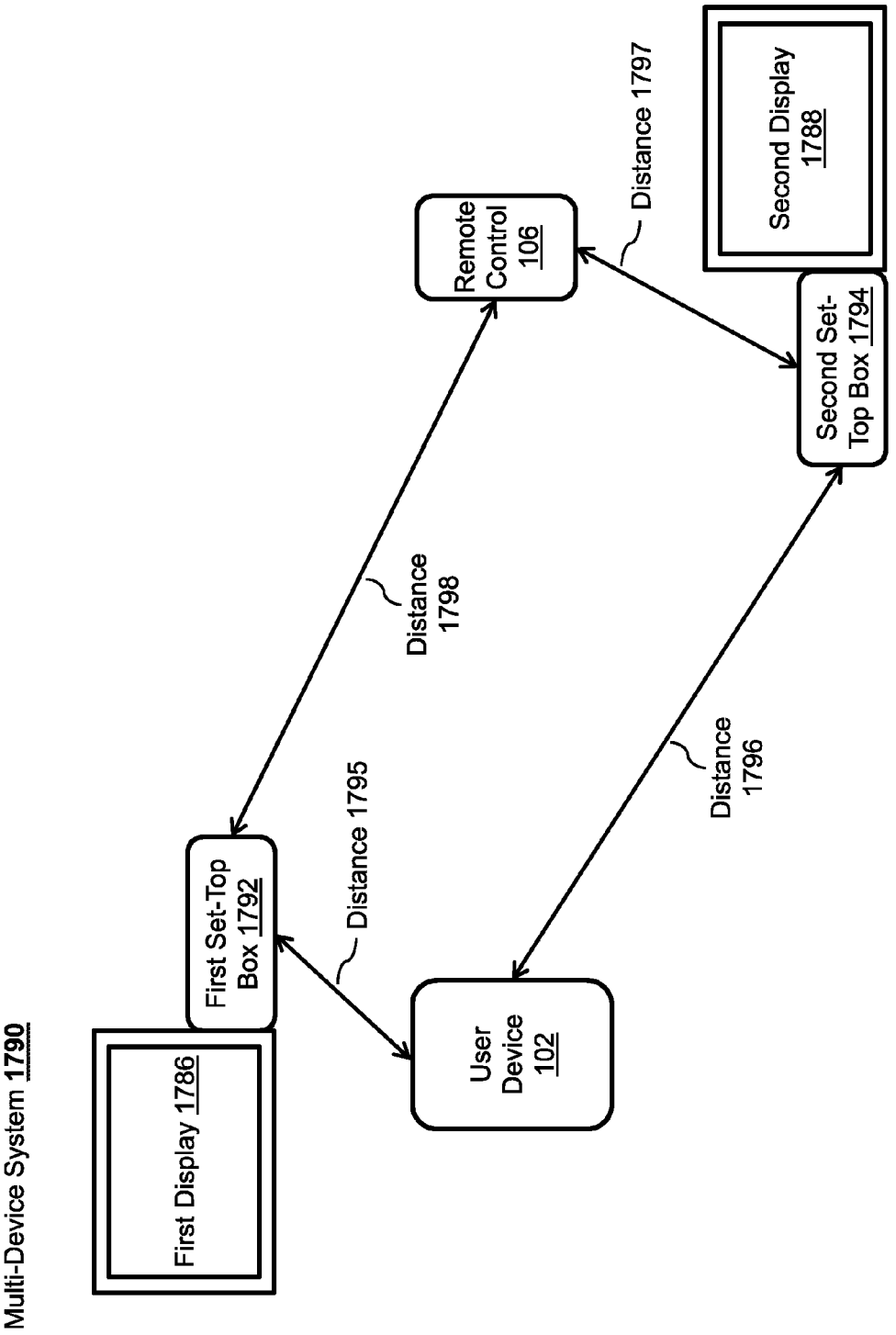

Multi-Device System 1790
First Display 1786
First Set-Top Box 1792
User Device 102
Remote Control 106
Second Set-Top Box 1794
Second Display 1788
Distance 1795
Distance 1796
Distance 1797
Distance 1798

FIG. 17

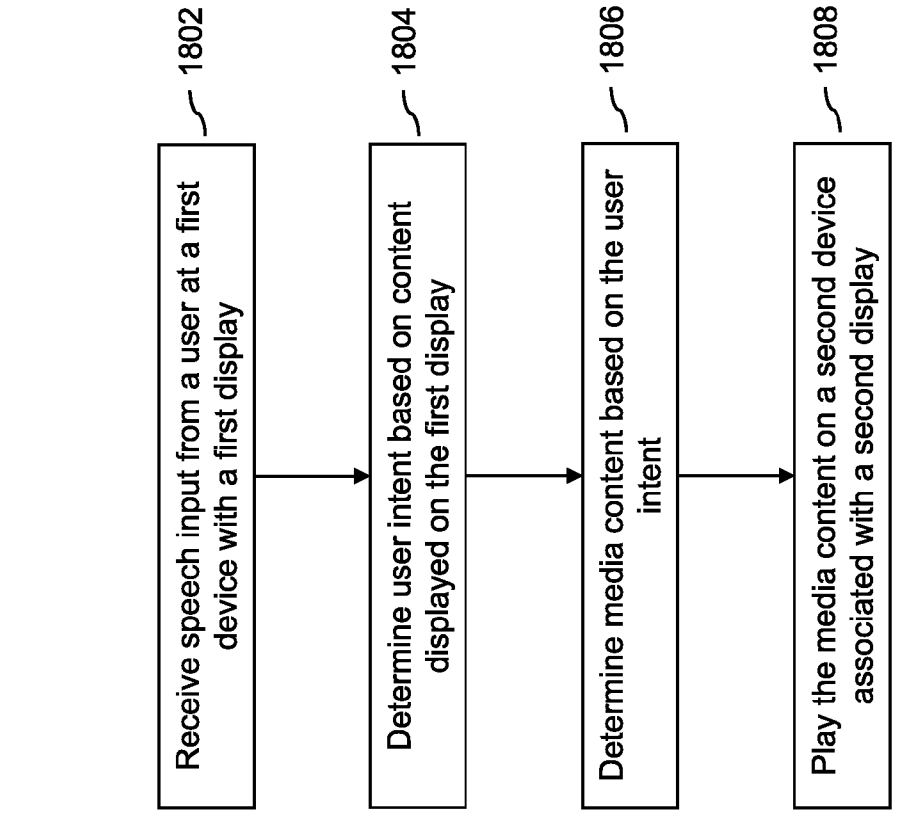

Process
1800
1802
Receive speech input from a user at a first device with a first display
1804
Determine user intent based on content displayed on the first display
1806
Determine media content based on the user intent
1808
Play the media content on a second device associated with a second display

FIG. 18

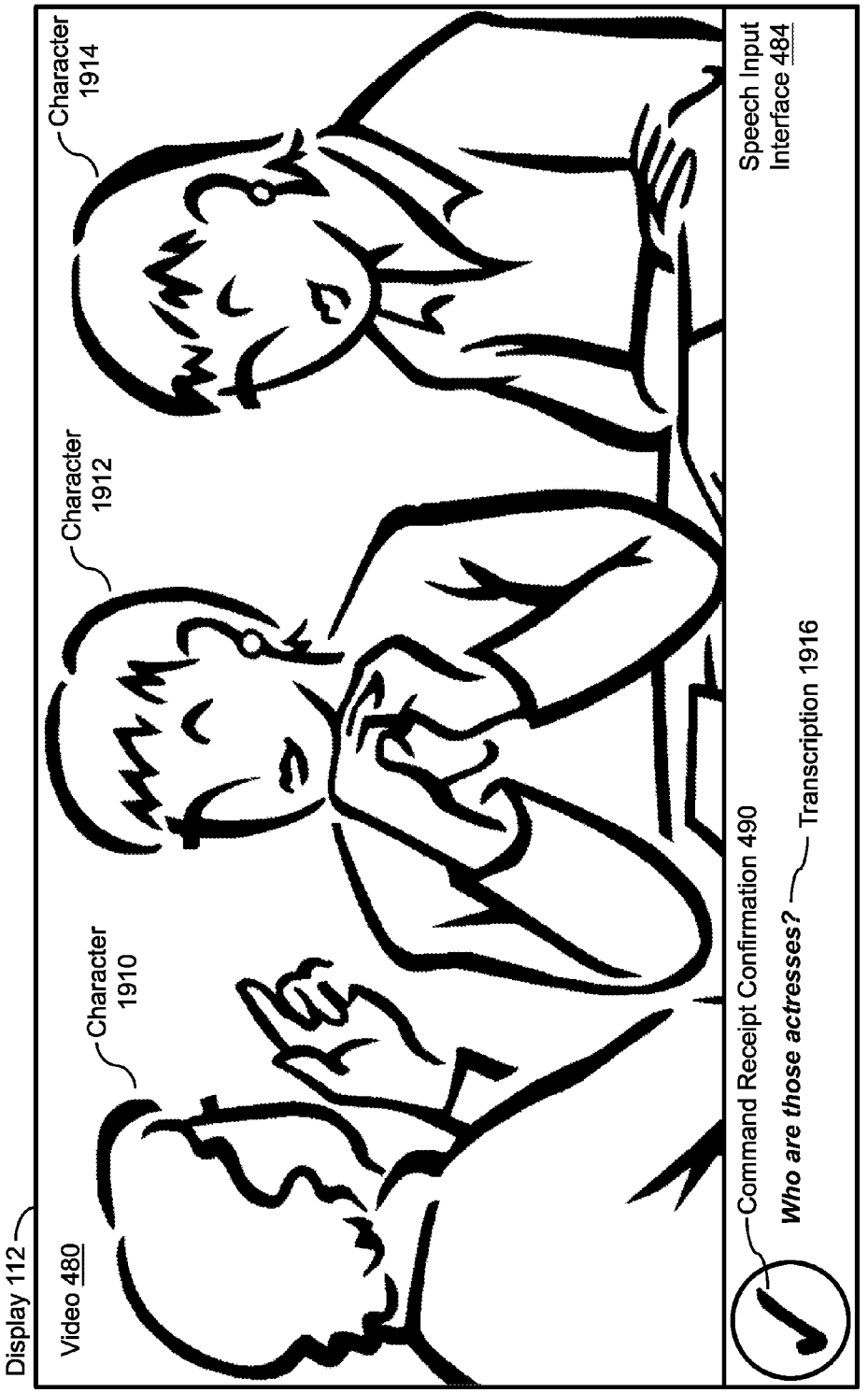

Display 112
Video 480
Character 1910
Character 1912
Character 1914
Speech Input Interface 484
Command Receipt Confirmation 490
Who are those actresses?
Transcription 1916

FIG. 19

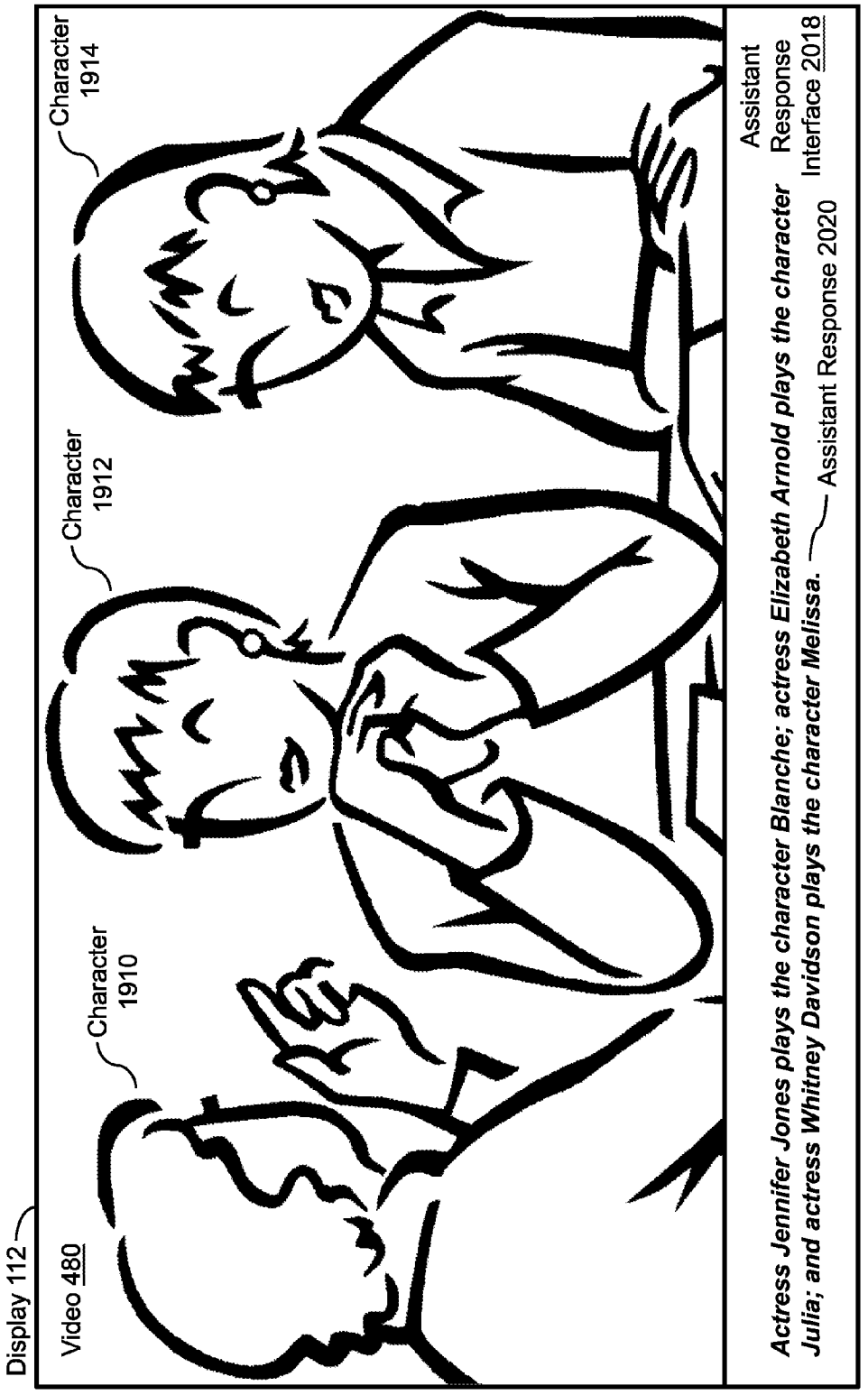

Display 112
Video 480
Character 1910
Character 1912
Character 1914
Actress Jennifer Jones plays the character Blanche; actress Elizabeth Arnold plays the character Julia; and actress Whitney Davidson plays the character Melissa.
Assistant Response 2020
Assistant Response Interface 2018

FIG. 20

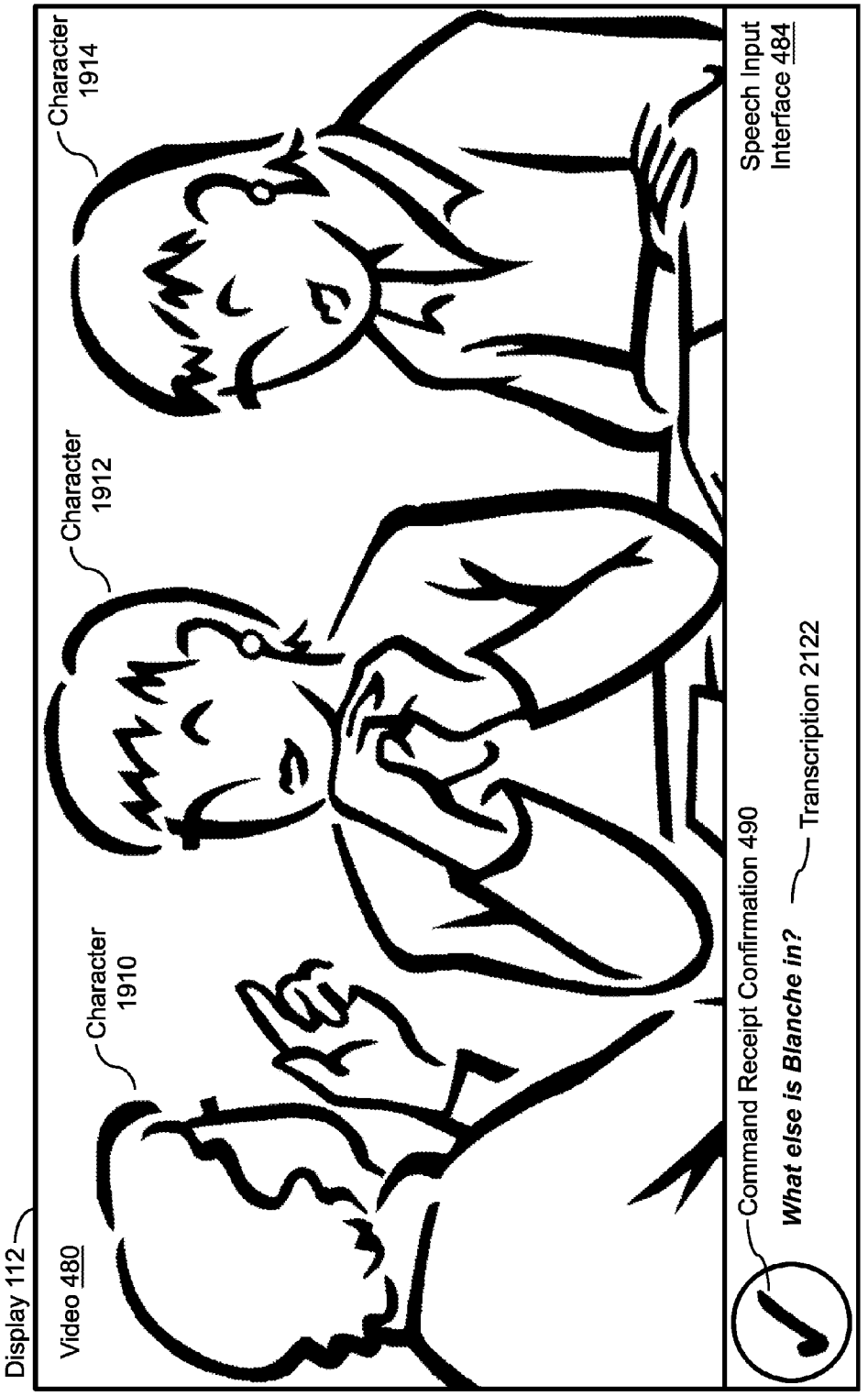

Display 112
Video 480
Character 1910
Character 1912
Character 1914
Speech Input Interface 484
Command Receipt Confirmation 490
What else is Blanche in?
Transcription 2122

FIG. 21

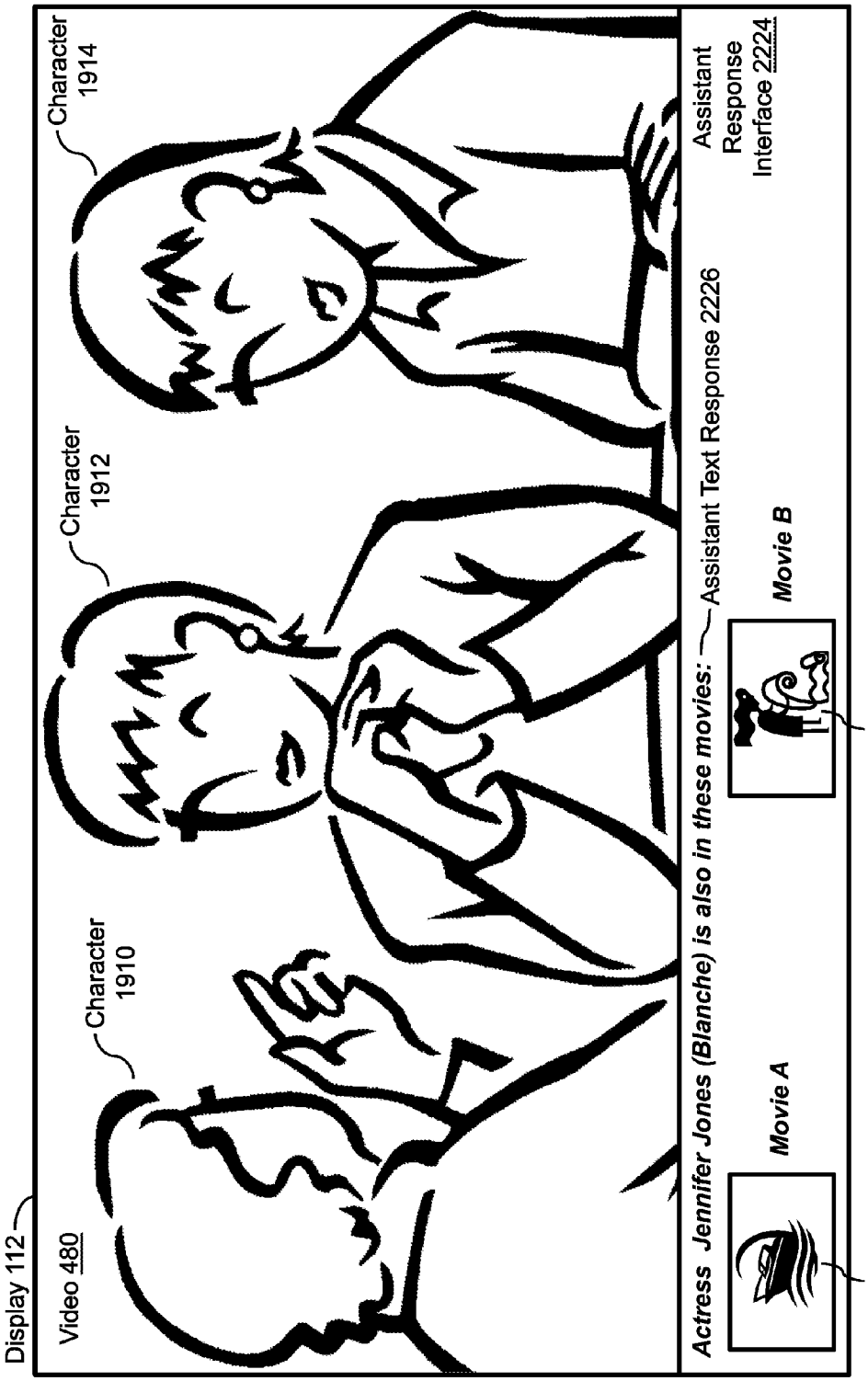

Display 112
Video 480
Character 1910
Character 1912
Character 1914
Actress Jennifer Jones (Blanche) is also in these movies:
Assistant Text Response 2226
Assistant Response Interface 2224
Movie A
Movie B
Selectable Video Link 2228
Selectable Video Link 2228

FIG. 22

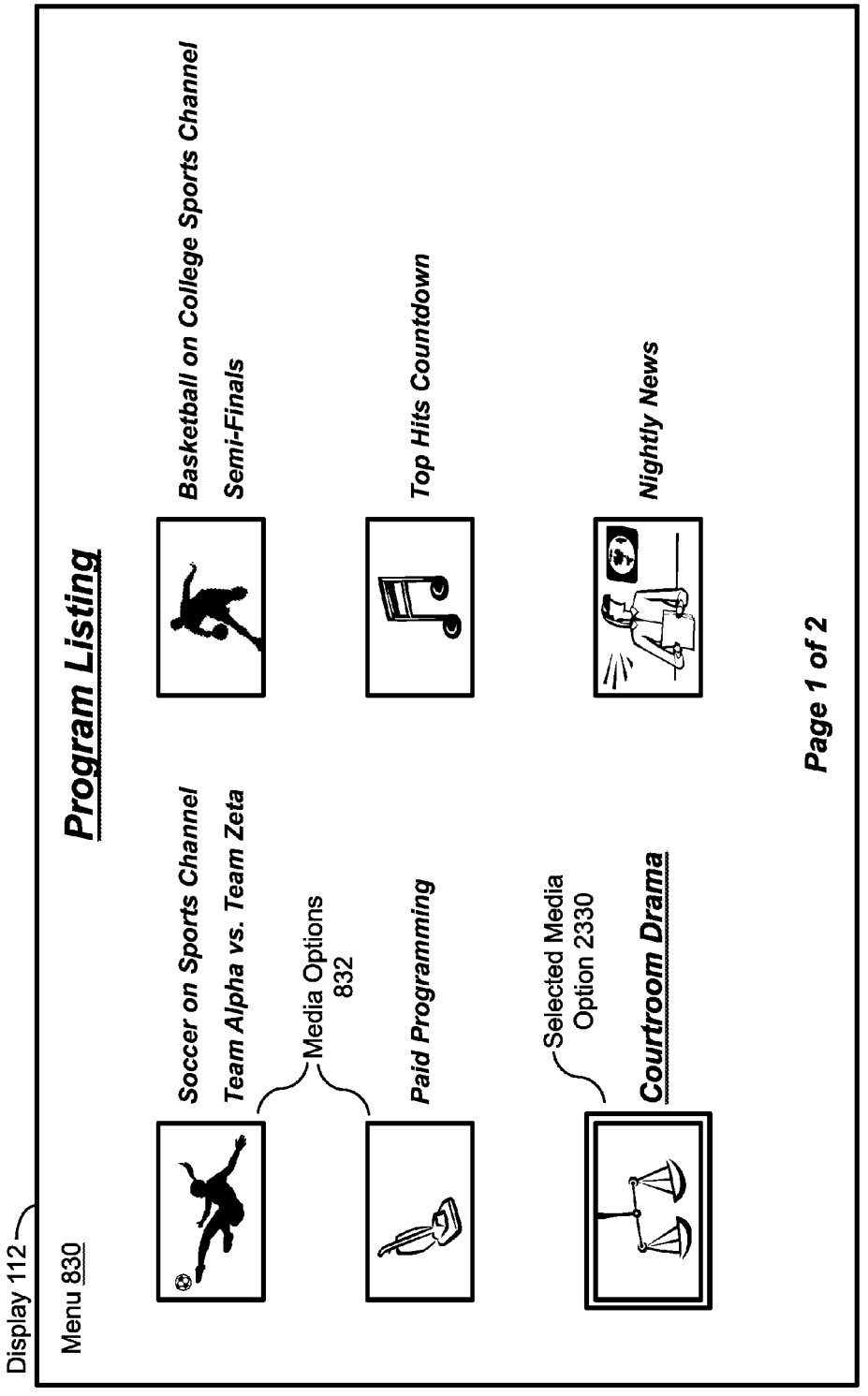

Display 112
Menu 830
Program Listing
Soccer on Sports Channel
Team Alpha vs. Team Zeta
Media Options 832
Paid Programming
Selected Media Option 2330
Courtroom Drama
Basketball on College Sports Channel
Semi-Finals
Top Hits Countdown
Nightly News
Page 1 of 2

FIG. 23A

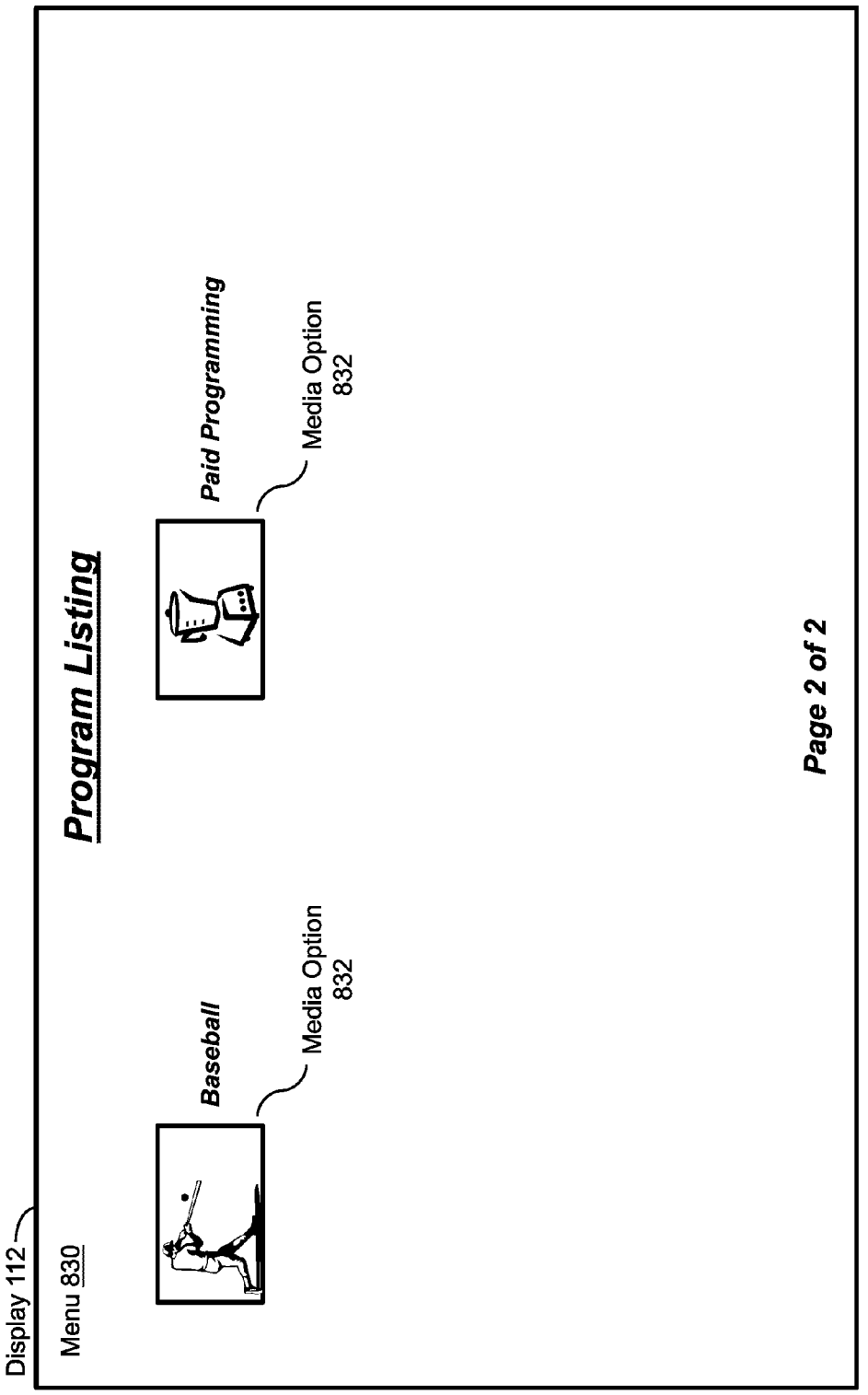

Display 112
Menu 830
Program Listing
Baseball
Media Option 832
Paid Programming
Media Option 832
Page 2 of 2

FIG. 23B

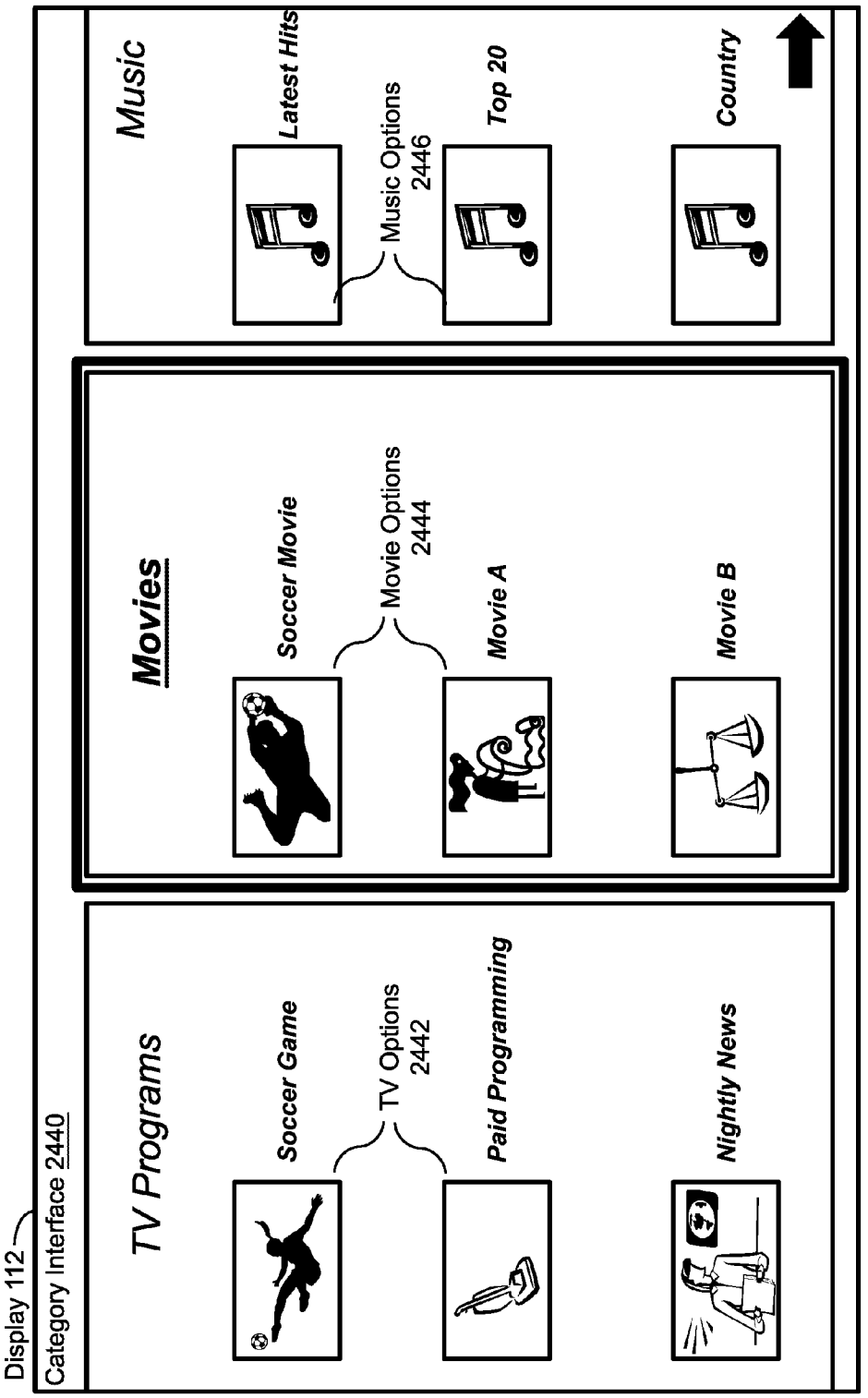

Display 112
Category Interface 2440
TV Programs
Soccer Game
TV Options 2442
Paid Programming
Nightly News
Movies
Soccer Movie
Movie Options 2444
Movie A
Movie B
Music
Latest Hits
Music Options 2446
Top 20
Country

FIG. 24

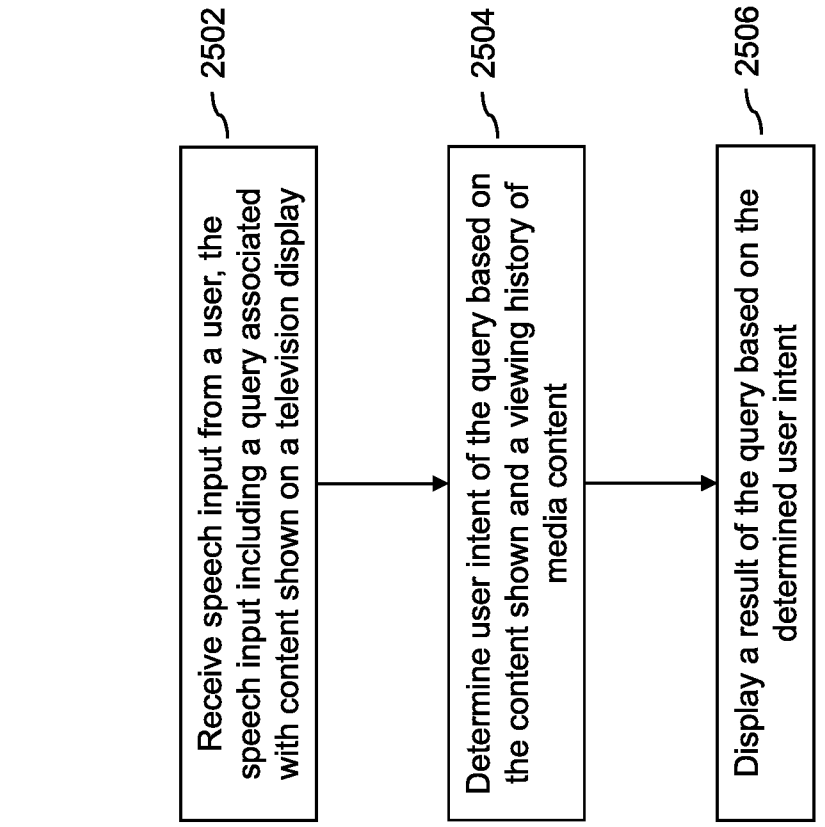

Process
2500
2502
Receive speech input from a user, the speech input including a query associated with content shown on a television display
2504
Determine user intent of the query based on the content shown and a viewing history of media content
2506
Display a result of the query based on the determined user intent

FIG. 25

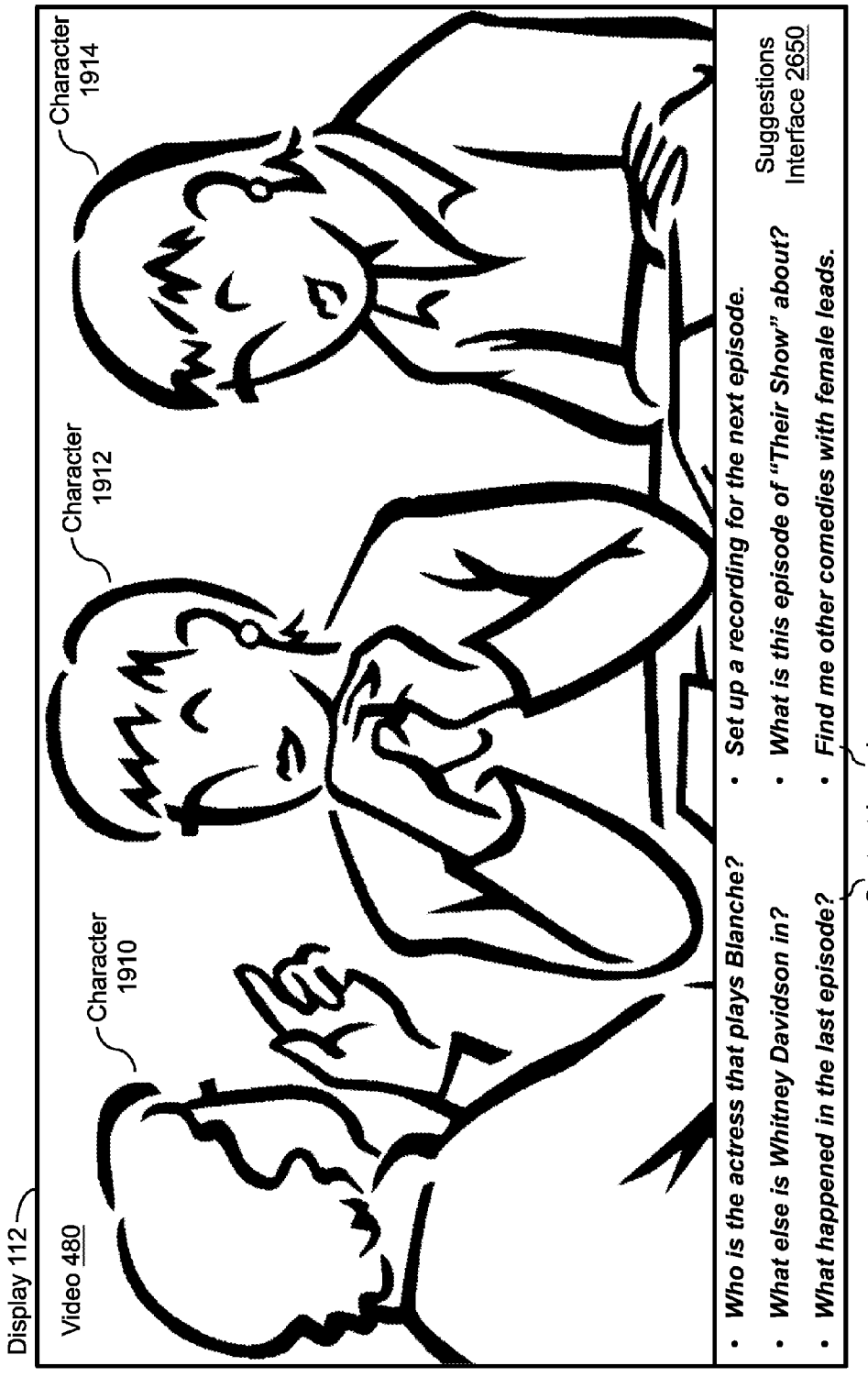

Display 112
Video 480
Character 1910
Character 1912
Character 1914
• Who is the actress that plays Blanche?
• What else is Whitney Davidson in?
• What happened in the last episode?
• Set up a recording for the next episode.
• What is this episode of "Their Show" about?
• Find me other comedies with female leads.
Content-based Suggestions 2652
Suggestions Interface 2650

FIG. 26

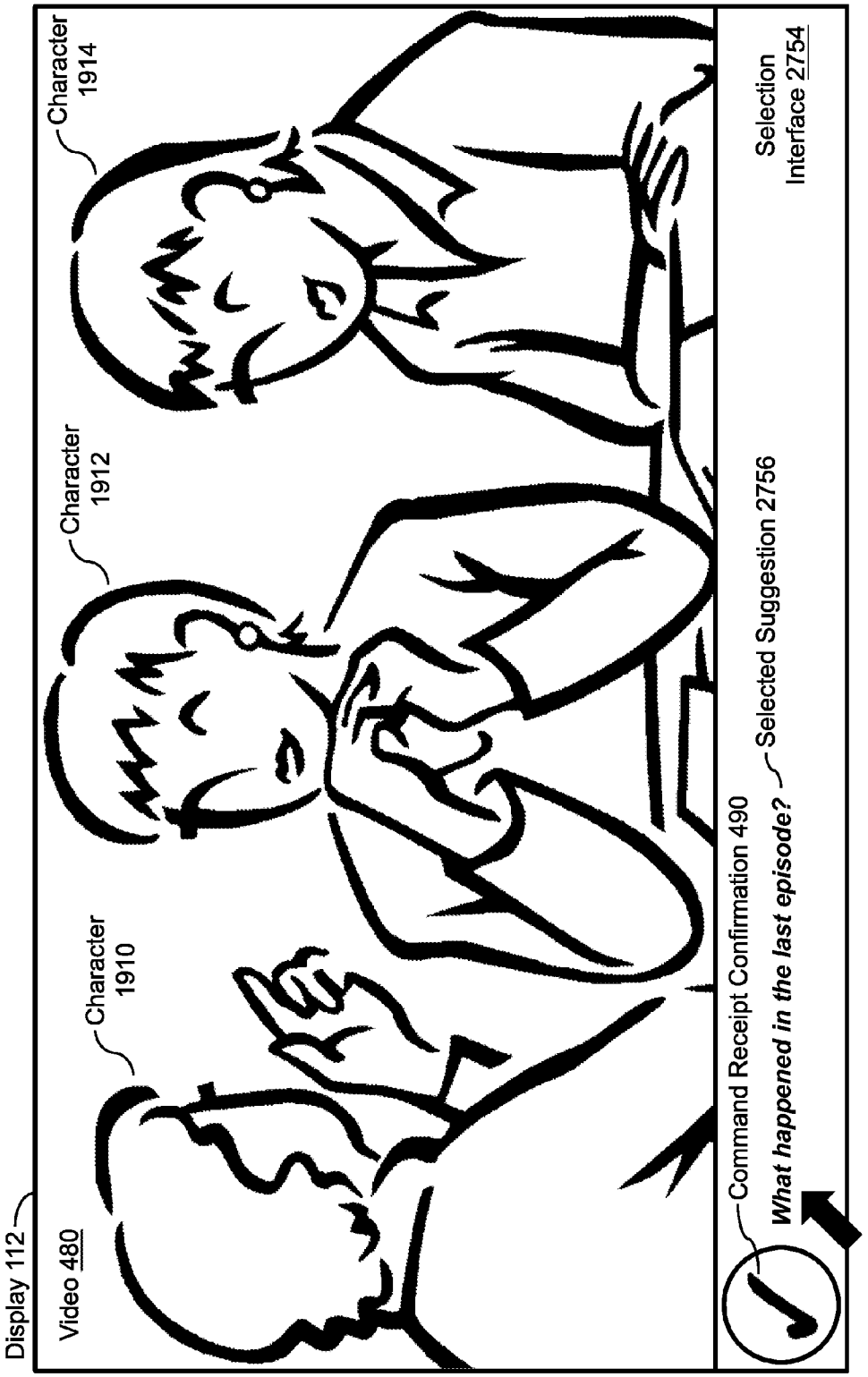

Display 112
Video 480
Character 1910
Character 1912
Character 1914
Command Receipt Confirmation 490
What happened in the last episode?
Selected Suggestion 2756
Selection Interface 2754

FIG. 27

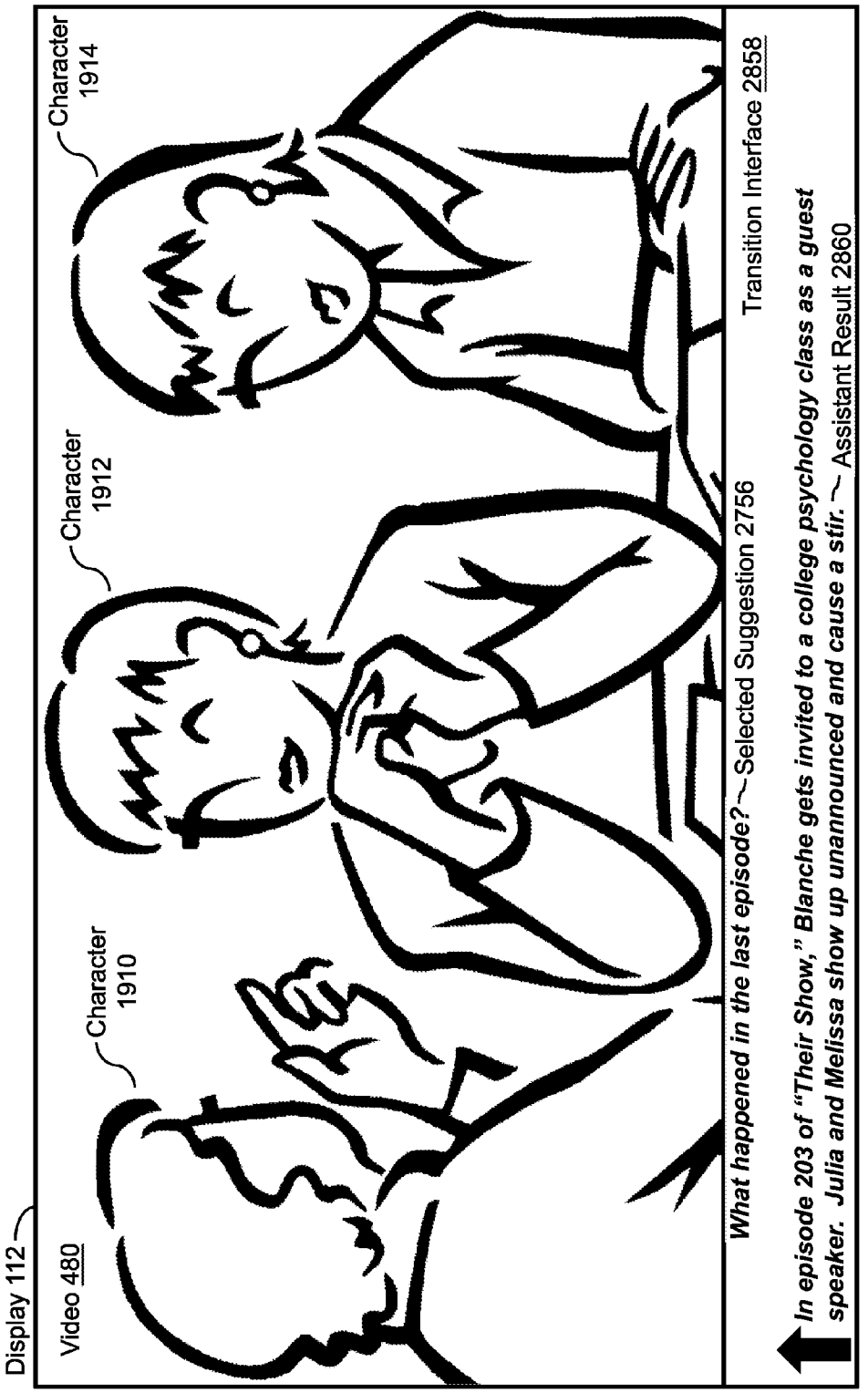

Display 112
Video 480
Character 1910
Character 1912
Character 1914
What happened in the last episode?
Selected Suggestion 2756
Transition Interface 2858
In episode 203 of "Their Show," Blanche gets invited to a college psychology class as a guest speaker. Julia and Melissa show up unannounced and cause a stir.
Assistant Result 2860

FIG. 28A

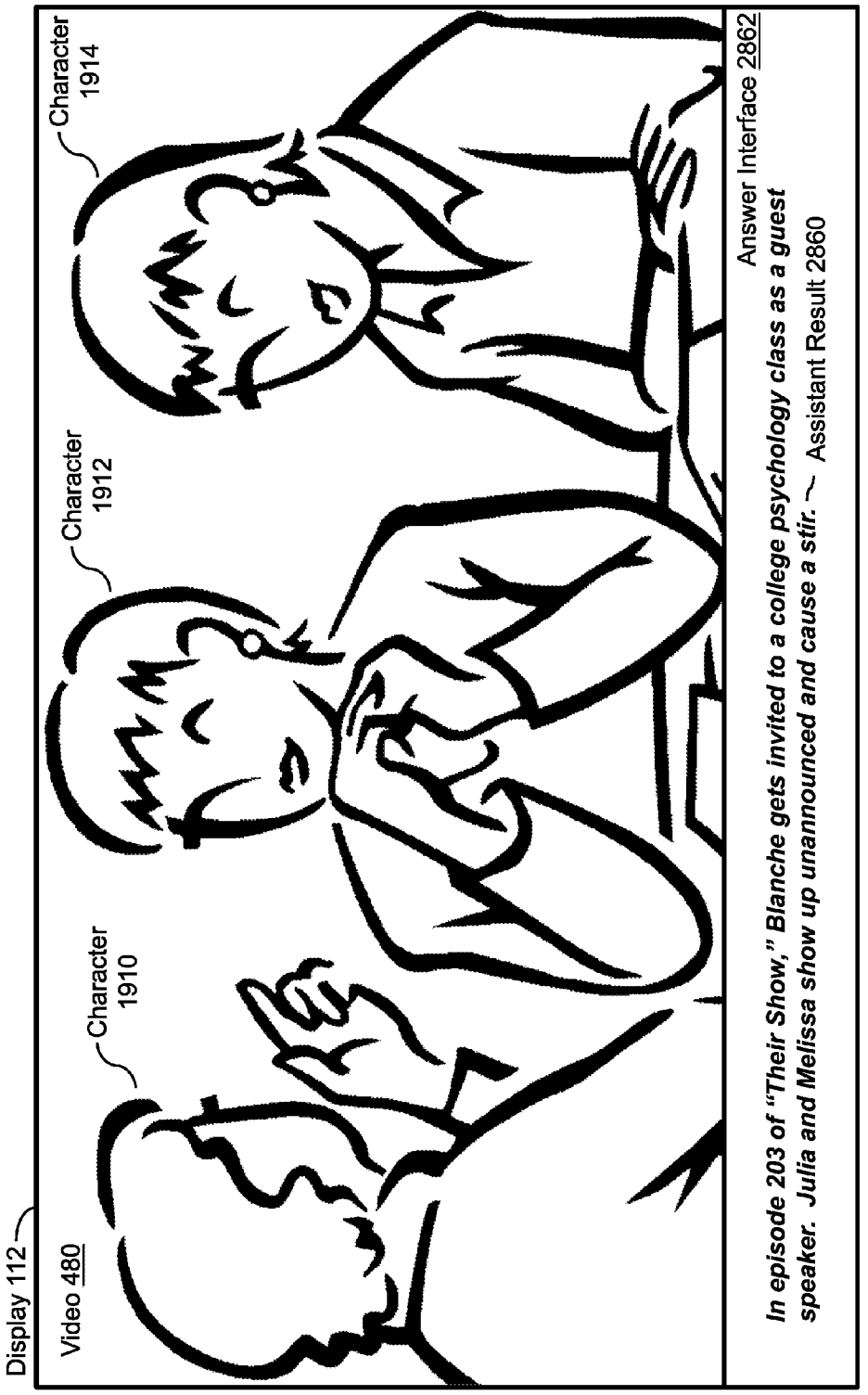

Display 112
Video 480
Character 1910
Character 1912
Character 1914
Answer Interface 2862
In episode 203 of "Their Show," Blanche gets invited to a college psychology class as a guest speaker. Julia and Melissa show up unannounced and cause a stir.
Assistant Result 2860

FIG. 28B

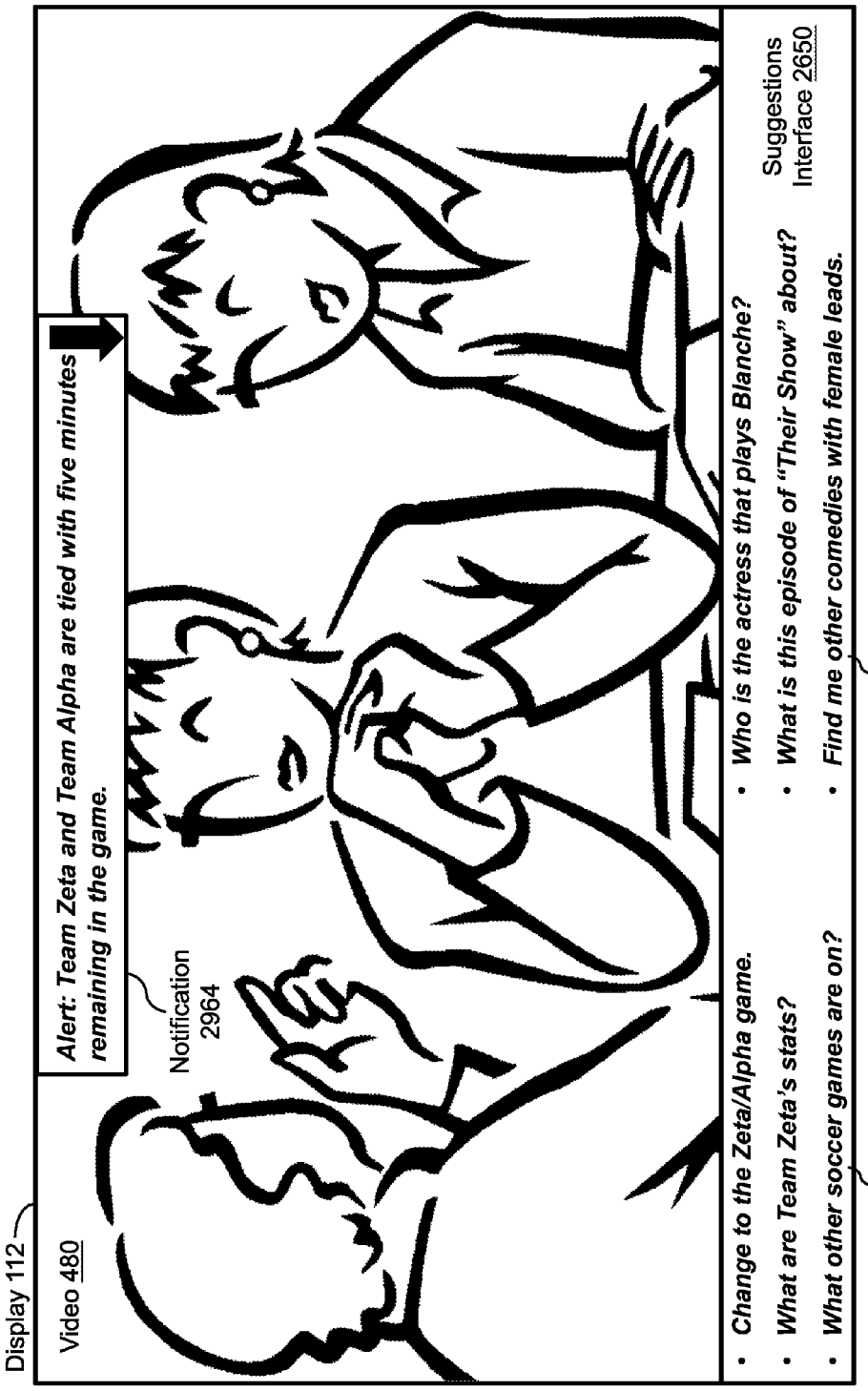

Display 112
Video 480
Alert: Team Zeta and Team Alpha are tied with five minutes remaining in the game.
Notification 2964
• Change to the Zeta/Alpha game.
• What are Team Zeta's stats?
• What other soccer games are on?
Notification-based Suggestions 2966
• Who is the actress that plays Blanche?
• What is this episode of "Their Show" about?
• Find me other comedies with female leads.
Content-based Suggestions 2652
Suggestions Interface 2650

FIG. 29

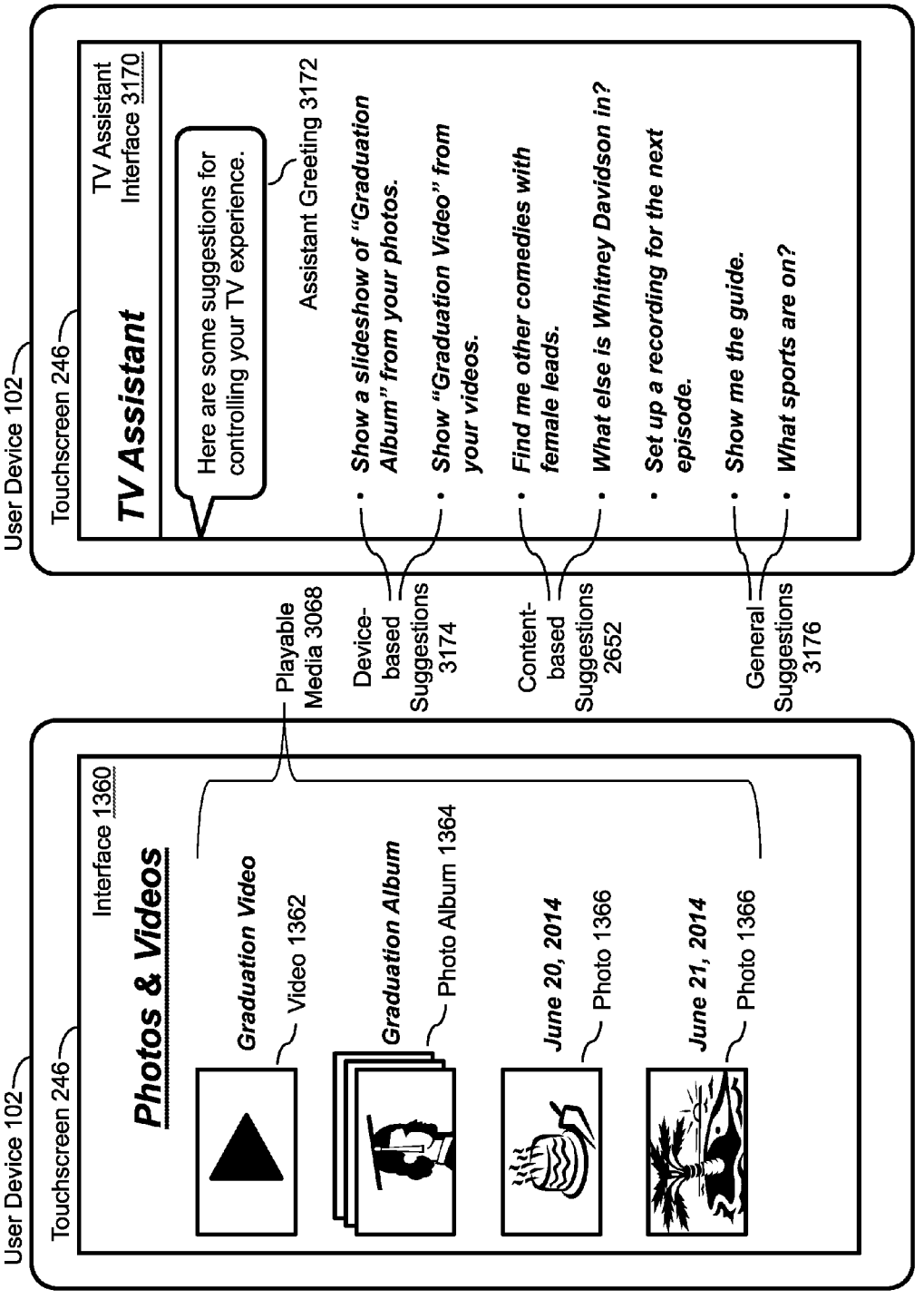

User Device 102
Touchscreen 246
Interface 1360
Photos & Videos
Graduation Video
Video 1362
Graduation Album
Photo Album 1364
June 20, 2014
Photo 1366
June 21, 2014
Photo 1366
Playable Media 3068
Device-based Suggestions 3174
Content-based Suggestions 2652
General Suggestions 3176
User Device 102
Touchscreen 246
TV Assistant
TV Assistant Interface 3170
Here are some suggestions for controlling your TV experience.
Assistant Greeting 3172
• Show a slideshow of "Graduation Album" from your photos.
• Show "Graduation Video" from your videos.
• Find me other comedies with female leads.
• What else is Whitney Davidson in?
• Set up a recording for the next episode.
• Show me the guide.
• What sports are on?

FIG. 30

FIG. 31

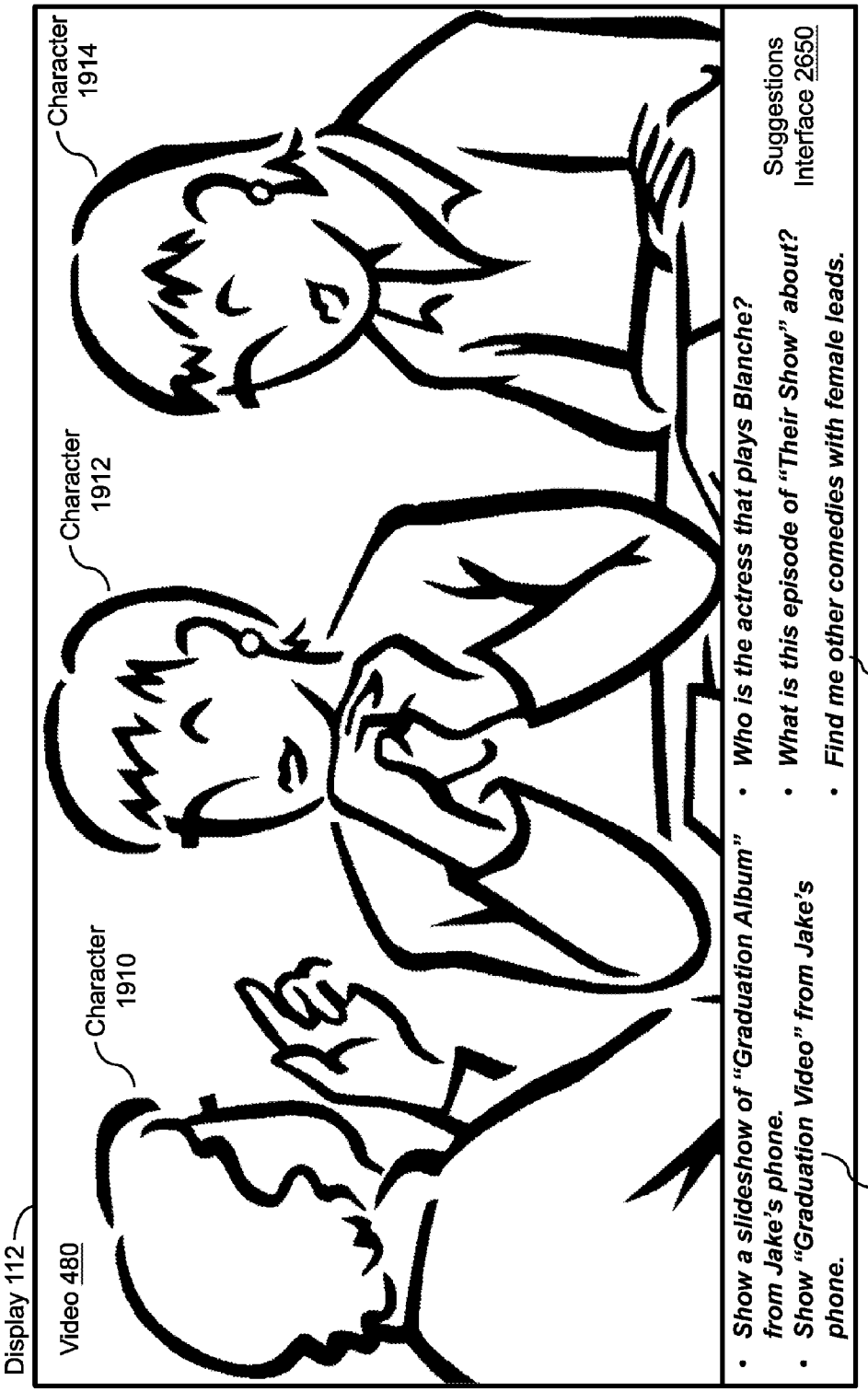

Display 112
Video 480
Character 1910
Character 1912
Character 1914
• Show a slideshow of "Graduation Album" from Jake's phone.
• Show "Graduation Video" from Jake's phone.
• Who is the actress that plays Blanche?
• What is this episode of "Their Show" about?
• Find me other comedies with female leads.
Suggestions Interface 2650
Connected Device-based Suggestions 3275
Content-based Suggestions 2652

FIG. 32

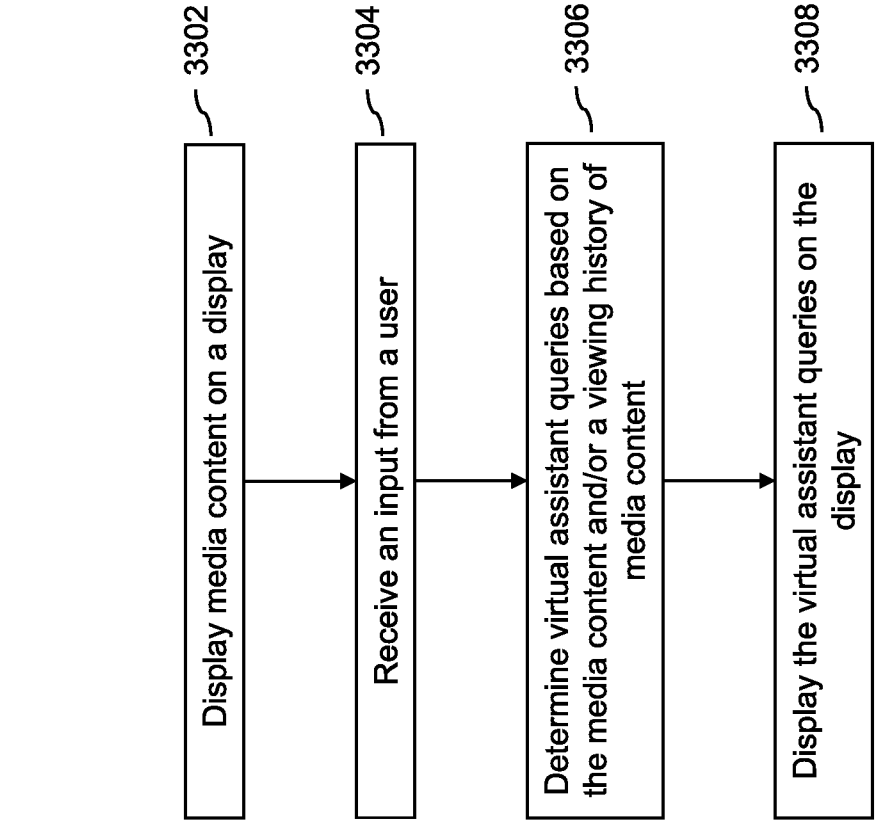

Process
3300
3302
Display media content on a display
3304
Receive an input from a user
3306
Determine virtual assistant queries based on the media content and/or a viewing history of media content
3308
Display the virtual assistant queries on the display

FIG. 33

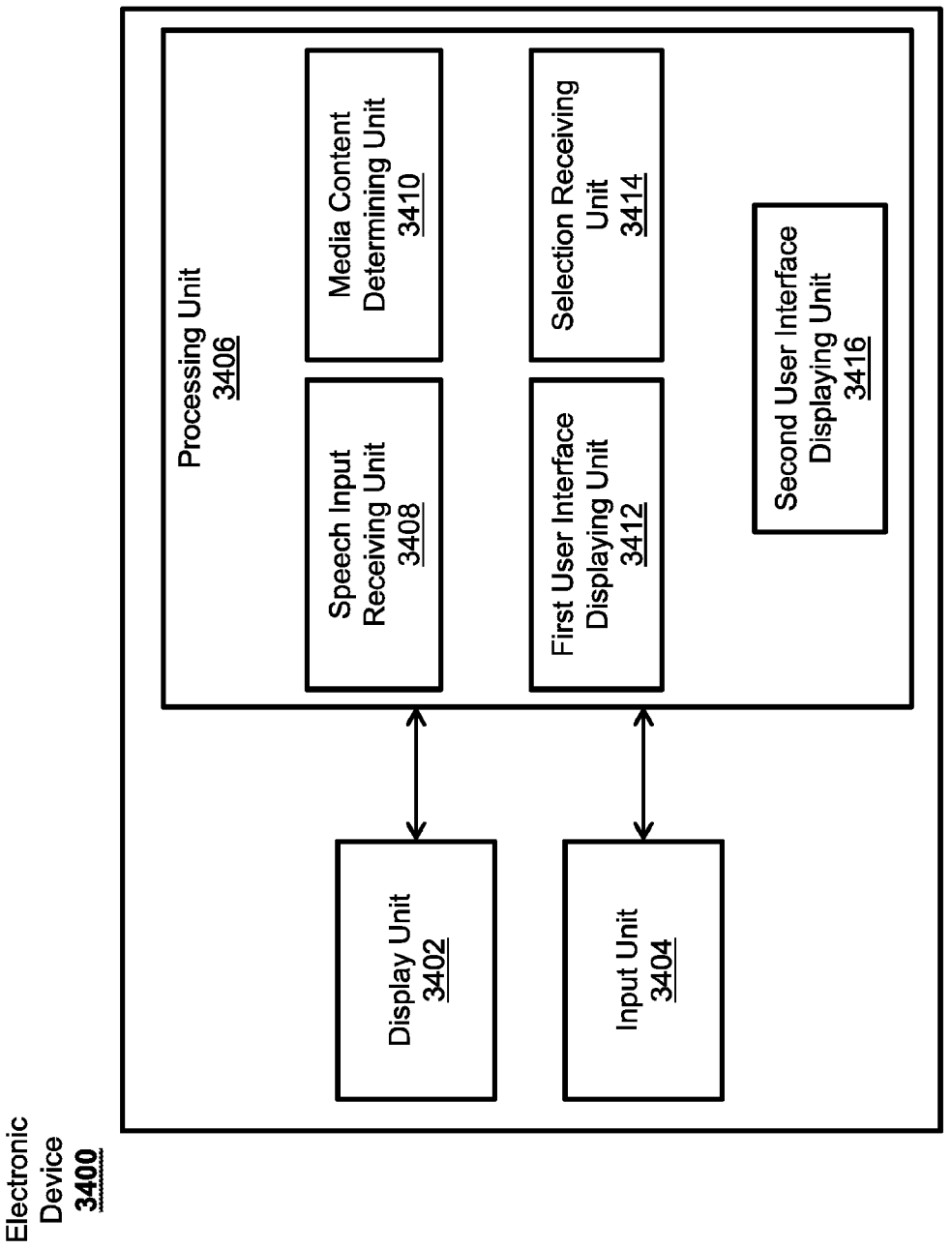

Electronic Device 3400
Processing Unit 3406
Media Content Determining Unit 3410
Selection Receiving Unit 3414
Speech Input Receiving Unit 3408
First User Interface Displaying Unit 3412
Second User Interface Displaying Unit 3416
Display Unit 3402
Input Unit 3404

FIG. 34

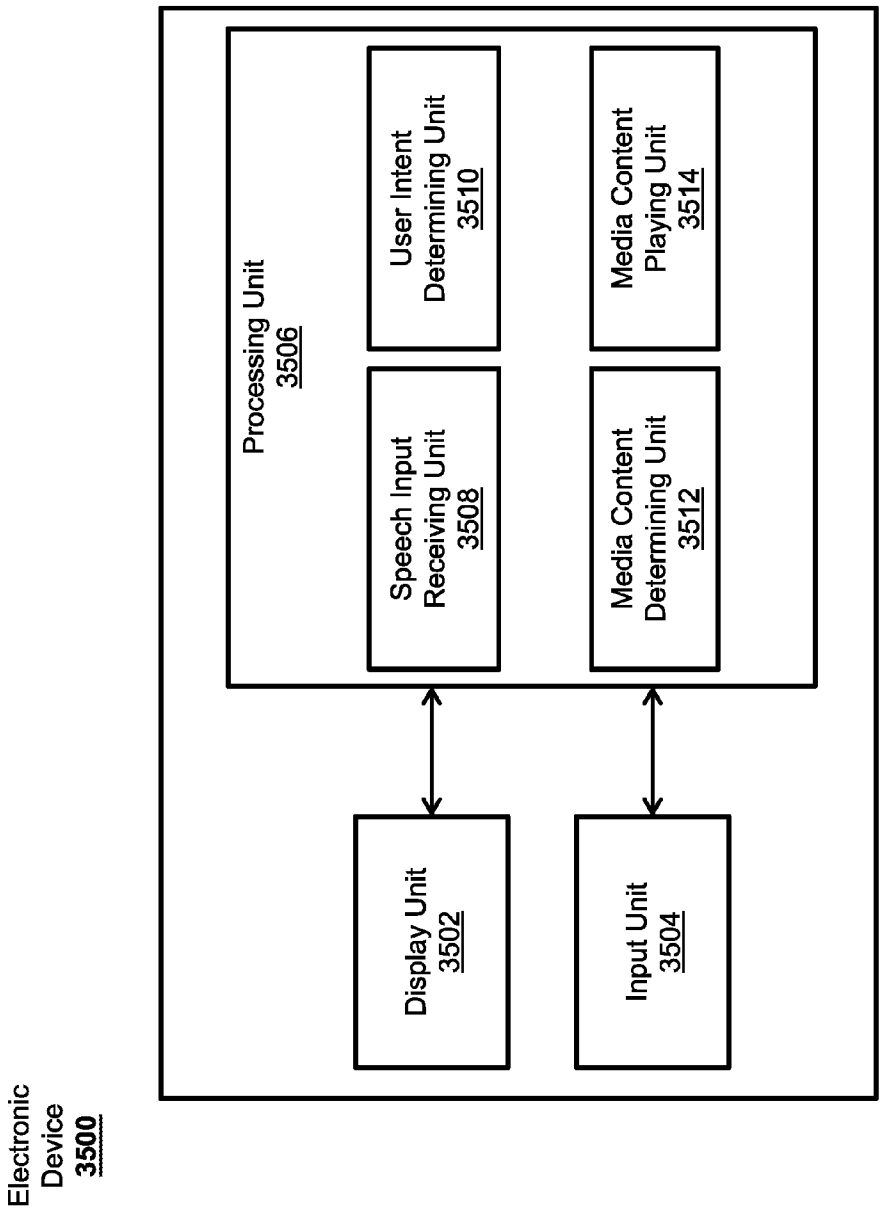

Electronic Device 3500
Processing Unit 3506
User Intent Determining Unit 3510
Media Content Playing Unit 3514
Speech Input Receiving Unit 3508
Media Content Determining Unit 3512
Display Unit 3502
Input Unit 3504

FIG. 35

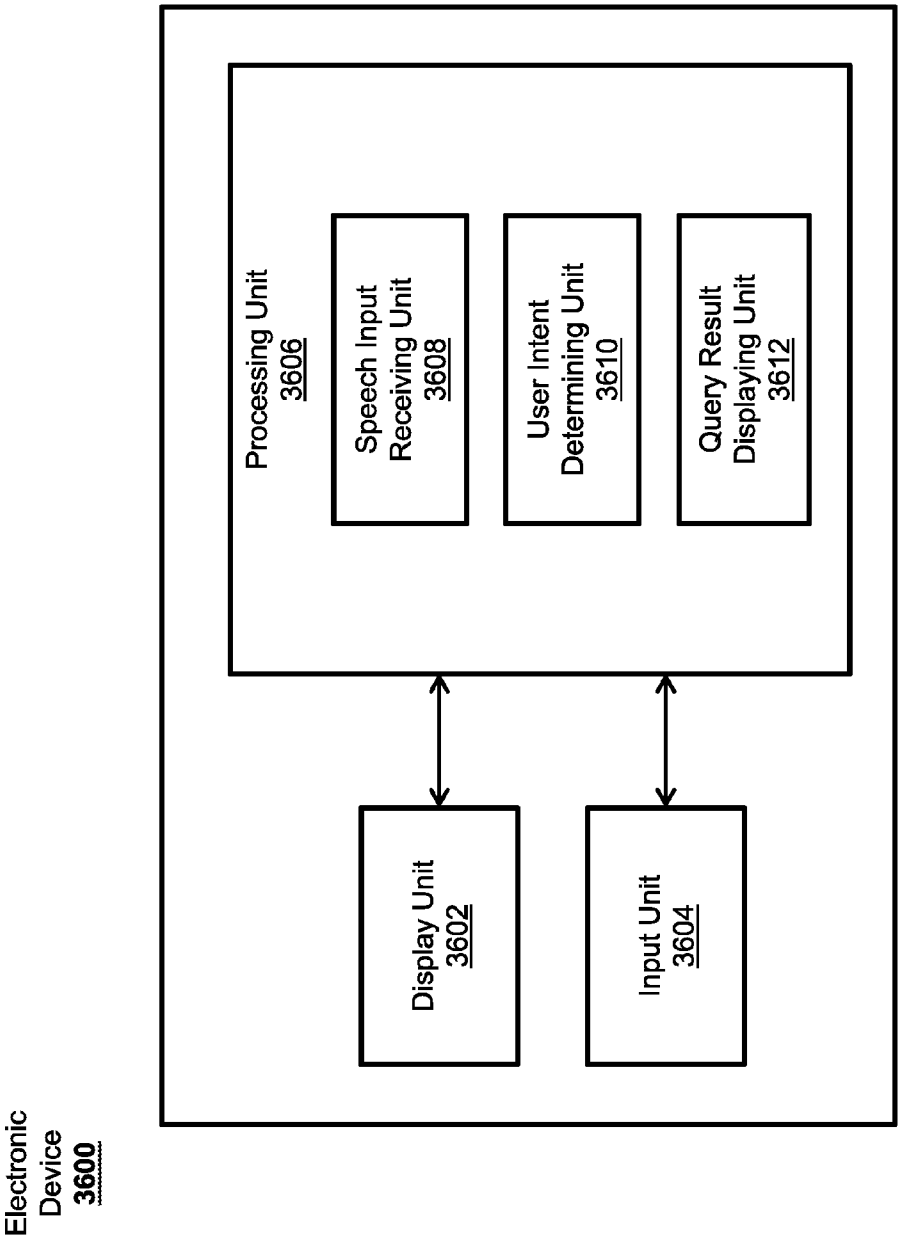

Electronic Device 3600
Processing Unit 3606
Speech Input Receiving Unit 3608
User Intent Determining Unit 3610
Query Result Displaying Unit 3612
Display Unit 3602
Input Unit 3604

FIG. 36

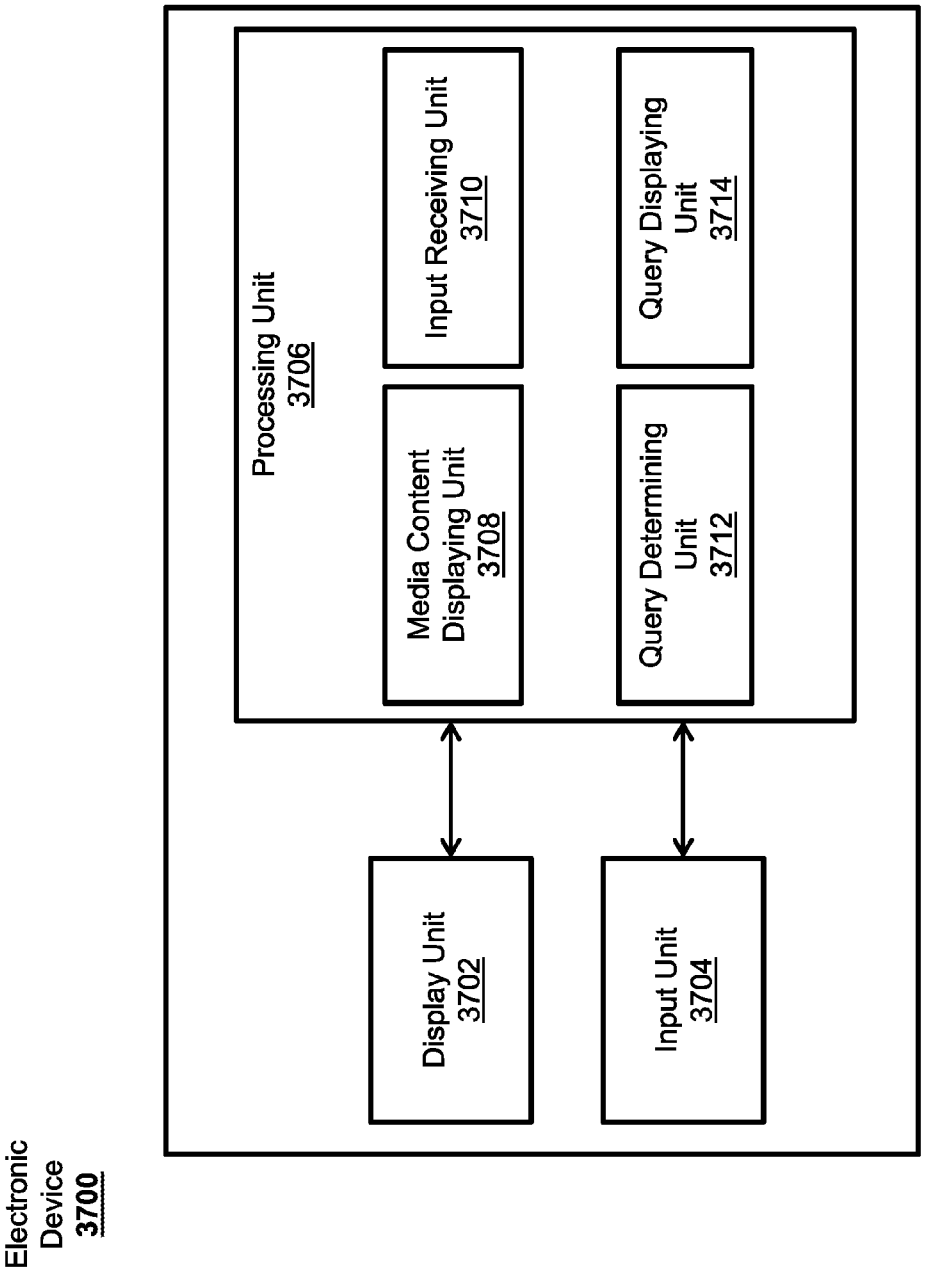

Electronic Device 3700
Processing Unit 3706
Input Receiving Unit 3710
Query Displaying Unit 3714
Media Content Displaying Unit 3708
Query Determining Unit 3712
Display Unit 3702
Input Unit 3704

FIG. 37

INTELLIGENT AUTOMATED ASSISTANT FOR TV USER INTERACTIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority from U.S. Provisional Ser. No. 62/019,312, filed on Jun. 30, 2014, entitled INTELLIGENT AUTOMATED ASSISTANT FOR TV USER INTERACTIONS, which is hereby incorporated by reference in its entirety for all purposes.

[0002] This application also relates to the following co-pending provisional application: U.S. Patent Application Ser. No. 62/019,292, "Real-time Digital Assistant Knowledge Updates," filed Jun. 30, 2014 (Attorney Docket No. 106843097900 (P22498USP1)), which is hereby incorporated by reference in its entirety.

FIELD

[0003] This relates generally to controlling television user interactions and, more specifically, to processing speech for a virtual assistant to control television user interactions.

BACKGROUND

[0004] Intelligent automated assistants (or virtual assistants) provide an intuitive interface between users and electronic devices. These assistants can allow users to interact with devices or systems using natural language in spoken and/or text forms. For example, a user can access the services of an electronic device by providing a spoken user input in natural language form to a virtual assistant associated with the electronic device. The virtual assistant can perform natural language processing on the spoken user input to infer the user's intent and operationalize the user's intent into tasks. The tasks can then be performed by executing one or more functions of the electronic device, and, in some examples, a relevant output can be returned to the user in natural language form.

[0005] While mobile telephones (e.g., smartphones), tablet computers, and the like have benefitted from virtual assistant control, many other user devices lack such convenient control mechanisms. For example, user interactions with media control devices (e.g., televisions, television set-top boxes, cable boxes, gaming devices, streaming media devices, digital video recorders, etc.) can be complicated and difficult to learn. Moreover, with the growing sources of media available through such devices (e.g., over-the-air TV, subscription TV service, streaming video services, cable on-demand video services, web-based video services, etc.), it can be cumbersome or even overwhelming for some users to find desired media content to consume. As a result, many media control devices can provide an inferior user experience that can be frustrating for many users.

SUMMARY

[0006] Systems and processes are disclosed for controlling television interactions using a virtual assistant. In one example, speech input can be received from a user. Media content can be determined based on the speech input. A first user interface having a first size can be displayed, and the first user interface can include selectable links to the media content. A selection of one of the selectable links can be received. In response to the selection, a second user interface can be displayed having a second size larger than the first size, and the second user interface can include the media content associated with the selection.

[0007] In another example, speech input can be received from a user at a first device having a first display. A user intent of the speech input can be determined based on content displayed on the first display. Media content can be determined based on the user intent. The media content can be played on a second device associated with a second display.

[0008] In another example, speech input can be received from a user, and the speech input can include a query associated with content shown on a television display. A user intent of the query can be determined based on the content shown on the television display and/or a viewing history of media content. A result of the query can be displayed based on the determined user intent.

[0009] In another example, media content can be displayed on a display. An input can be received from a user. Virtual assistant queries can be determined based on the media content and/or a viewing history of media content. The virtual assistant queries can be displayed on the display.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] FIG. **1** illustrates an exemplary system for controlling television user interaction using a virtual assistant.

[0011] FIG. **2** illustrates a block diagram of an exemplary user device according to various examples.

[0012] FIG. **3** illustrates a block diagram of an exemplary media control device in a system for controlling television user interaction.

[0013] FIGS. **4**A-**4**E illustrate an exemplary speech input interface over video content.

[0014] FIG. **5** illustrates an exemplary media content interface over video content.

[0015] FIGS. **6**A-**6**B illustrate an exemplary media detail interface over video content.

[0016] FIGS. **7**A-**7**B illustrate an exemplary media transition interface.

[0017] FIGS. **8**A-**8**B illustrate an exemplary speech input interface over menu content.

[0018] FIG. **9** illustrates an exemplary virtual assistant result interface over menu content.

[0019] FIG. **10** illustrates an exemplary process for controlling television interactions using a virtual assistant and displaying associated information using different interfaces.

[0020] FIG. **11** illustrates exemplary television media content on a mobile user device.

[0021] FIG. **12** illustrates exemplary television control using a virtual assistant.

[0022] FIG. **13** illustrates exemplary picture and video content on a mobile user device.

[0023] FIG. **14** illustrates exemplary media display control using a virtual assistant.

[0024] FIG. **15** illustrates exemplary virtual assistant interactions with results on a mobile user device and a media display device.

[0025] FIG. **16** illustrates exemplary virtual assistant interactions with media results on a media display device and a mobile user device.

[0026] FIG. **17** illustrates exemplary media device control based on proximity.

[0027] FIG. **18** illustrates an exemplary process for controlling television interactions using a virtual assistant and multiple user devices.

[0028] FIG. **19** illustrates an exemplary speech input interface with a virtual assistant query about background video content.

[0029] FIG. **20** illustrates an exemplary informational virtual assistant response over video content.

[0030] FIG. **21** illustrates an exemplary speech input interface with a virtual assistant query for media content associated with background video content.

[0031] FIG. **22** illustrates an exemplary virtual assistant response interface with selectable media content.

[0032] FIGS. **23**A-**23**B illustrate exemplary pages of a program menu.

[0033] FIG. **24** illustrates an exemplary media menu divided into categories.

[0034] FIG. **25** illustrates an exemplary process for controlling television interactions using media content shown on a display and a viewing history of media content.

[0035] FIG. **26** illustrates an exemplary interface with virtual assistant query suggestions based on background video content.

[0036] FIG. **27** illustrates an exemplary interface for confirming selection of a suggested query.

[0037] FIGS. **28**A-**28**B illustrate an exemplary virtual assistant answer interface based on a selected query.

[0038] FIG. **29** illustrates a media content notification and an exemplary interface with virtual assistant query suggestions based on the notification.

[0039] FIG. **30** illustrates a mobile user device with exemplary picture and video content that is playable on a media control device.

[0040] FIG. **31** illustrates an exemplary mobile user device interface with virtual assistant query suggestions based on playable user device content and based on video content shown on a separate display

[0041] FIG. **32** illustrates an exemplary interface with virtual assistant query suggestions based on playable content from a separate user device.

[0042] FIG. **33** illustrates an exemplary process for suggesting virtual assistant interactions for controlling media content.

[0043] FIG. **34** illustrates a functional block diagram of an electronic device configured to control television interactions using a virtual assistant and display associated information using different interfaces according to various examples.

[0044] FIG. **35** illustrates a functional block diagram of an electronic device configured to control television interactions using a virtual assistant and multiple user devices according to various examples.

[0045] FIG. **36** illustrates a functional block diagram of an electronic device configured to control television interactions using media content shown on a display and a viewing history of media content according to various examples.

[0046] FIG. **37** illustrates a functional block diagram of an electronic device configured to suggest virtual assistant interactions for controlling media content according to various examples.

DETAILED DESCRIPTION

[0047] In the following description of examples, reference is made to the accompanying drawings in which it is shown by way of illustration specific examples that can be practiced. It is to be understood that other examples can be used and structural changes can be made without departing from the scope of the various examples.

[0048] This relates to systems and processes for controlling television user interactions using a virtual assistant. In one example, a virtual assistant can be used to interact with a media control device, such as a television set-top box controlling content shown on a television display. A mobile user device or a remote control with a microphone can be used to receive speech input for the virtual assistant. The user's intent can be determined from the speech input, and the virtual assistant can execute tasks according to the user's intent, including causing playback of media on a connected television and controlling any other functions of a television set-top box or like device (e.g., managing video recordings, searching for media content, navigating menus, etc.).

[0049] Virtual assistant interactions can be shown on a connected television or other display. In one example, media content can be determined based on speech input received from a user. A first user interface with a first small size can be displayed, including selectable links to the determined media content. After receiving a selection of a media link, a second user interface with a second larger size can be displayed, including the media content associated with the selection. In other examples, the interface used to convey virtual assistant interactions can expand or contract to occupy a minimal amount of space while conveying desired information.

[0050] In some examples, multiple devices associated with multiple displays can be used to determine user intent from speech input as well as to convey information to users in different ways. For example, speech input can be received from a user at a first device having a first display. The user's intent can be determined from the speech input based on content displayed on the first display. Media content can be determined based on the user intent, and the media content can be played on a second device associated with a second display.

[0051] Television display content can also be used as contextual input for determining user intent from speech input. For example, speech input can be received from a user, including a query associated with content shown on a television display. The user intent of the query can be determined based on the content shown on the television display as well as a viewing history of media content on the television display (e.g., disambiguating the query based on characters in a playing TV show). The results of the query can then be displayed based on the determined user intent.

[0052] In some examples, virtual assistant query suggestions can be provided to the user (e.g., to acquaint the user with available commands, suggest interesting content, etc.). For example, media content can be shown on a display, and an input can be received from the user requesting virtual assistant query suggestions. Virtual assistant queries suggestions can be determined based on the media content shown on the display and a viewing history of media content shown on the display (e.g., suggesting queries related to a playing TV show). The suggested virtual assistant queries can then be shown on the display.

[0053] Controlling television user interactions using a virtual assistant according to the various examples discussed herein can provide an efficient and enjoyable user experience. User interactions with media control devices can be intuitive and simple using a virtual assistant capable of receiving natural language queries or commands. Available functions can be suggested to users as desired, including meaningful query suggestions based on playing content, which can aid users to learn control capabilities. In addition, available media can be

made easily accessible using intuitive spoken commands. It should be understood, however, that still many other advantages can be achieved according to the various examples discussed herein.

[0054] FIG. **1** illustrates exemplary system **100** for controlling television user interaction using a virtual assistant. It should be understood that controlling television user interaction as discussed herein is merely one example of controlling media on one type of display technology and is used for reference, and the concepts discussed herein can be used for controlling any media content interactions generally, including on any of a variety of devices and associated displays (e.g., monitors, laptop displays, desktop computer displays, mobile user device displays, projector displays, etc.). The term "television" can thus refer to any type of display associated with any of a variety of devices. Moreover, the terms "virtual assistant," "digital assistant," "intelligent automated assistant," or "automatic digital assistant" can refer to any information processing system that can interpret natural language input in spoken and/or textual form to infer user intent, and perform actions based on the inferred user intent. For example, to act on an inferred user intent, the system can perform one or more of the following: identifying a task flow with steps and parameters designed to accomplish the inferred user intent; inputting specific requirements from the inferred user intent into the task flow; executing the task flow by invoking programs, methods, services, APIs, or the like; and generating output responses to the user in an audible (e.g., spoken) and/or visual form.

[0055] A virtual assistant can be capable of accepting a user request at least partially in the form of a natural language command, request, statement, narrative, and/or inquiry. Typically, the user request seeks either an informational answer or performance of a task by the virtual assistant (e.g., causing display of particular media). A satisfactory response to the user request can include provision of the requested informational answer, performance of the requested task, or a combination of the two. For example, a user can ask the virtual assistant a question, such as "Where am I right now?" Based on the user's current location, the virtual assistant can answer, "You are in Central Park." The user can also request the performance of a task, for example, "Please remind me to call Mom at 4 p.m. today." In response, the virtual assistant can acknowledge the request and then create an appropriate reminder item in the user's electronic schedule. During the performance of a requested task, the virtual assistant can sometimes interact with the user in a continuous dialogue involving multiple exchanges of information over an extended period of time. There are numerous other ways of interacting with a virtual assistant to request information or performance of various tasks. In addition to providing verbal responses and taking programmed actions, the virtual assistant can also provide responses in other visual or audio forms (e.g., as text, alerts, music, videos, animations, etc.). Moreover, as discussed herein, an exemplary virtual assistant can control playback of media content (e.g., playing video on a television) and cause information to be displayed on a display.

[0056] An example of a virtual assistant is described in Applicants' U.S. Utility application Ser. No. 12/987,982 for "Intelligent Automated Assistant," filed Jan. 10, 2011, the entire disclosure of which is incorporated herein by reference.

[0057] As shown in FIG. **1**, in some examples, a virtual assistant can be implemented according to a client-server model. The virtual assistant can include a client-side portion executed on a user device **102** and a server-side portion executed on a server system **110**. The client-side portion can also be executed on television set-top box **104** in conjunction with remote control **106**. User device **102** can include any electronic device, such as a mobile phone (e.g., smartphone), tablet computer, portable media player, desktop computer, laptop computer, PDA, wearable electronic device (e.g., digital glasses, wristband, wristwatch, brooch, armband, etc.), or the like. Television set-top box **104** can include any media control device, such as a cable box, satellite box, video player, video streaming device, digital video recorder, gaming system, DVD player, Blu-ray Disc™ Player, a combination of such devices, or the like. Television set-top box **104** can be connected to display **112** and speakers **111** via a wired or wireless connection. Display **112** (with or without speakers **111**) can be any type of display, such as a television display, monitor, projector, or the like. In some examples, television set-top box **104** can connect to an audio system (e.g., audio receiver), and speakers **111** can be separate from display **112**. In other examples, display **112**, speakers **111**, and television set-top box **104** can be incorporated together in a single device, such as a smart television with advanced processing and network connectivity capabilities. In such examples, the functions of television set-top box **104** can be executed as an application on the combined device.

[0058] In some examples, television set-top box **104** can function as a media control center for multiple types and sources of media content. For example, television set-top box **104** can facilitate user access to live television (e.g., over-the-air, satellite, or cable television). As such, television set-top box **104** can include cable tuners, satellite tuners, or the like. In some examples, television set-top box **104** can also record television programs for later time-shifted viewing. In other examples, television set-top box **104** can provide access to one or more streaming media services, such as cable-delivered on-demand television shows, videos, and music as well as Internet-delivered television shows, videos, and music (e.g., from various free, paid, and subscription-based streaming services). In still other examples, television set-top box **104** can facilitate playback or display of media content from any other source, such as displaying photos from a mobile user device, playing videos from a coupled storage device, playing music from a coupled music player, or the like. Television set-top box **104** can also include various other combinations of the media control features discussed herein, as desired.

[0059] User device **102** and television set-top box **104** can communicate with server system **110** through one or more networks **108**, which can include the Internet, an intranet, or any other wired or wireless public or private network. In addition, user device **102** can communicate with television set-top box **104** through network **108** or directly through any other wired or wireless communication mechanisms (e.g., Bluetooth, Wi-Fi, radio frequency, infrared transmission, etc.). As illustrated, remote control **106** can communicate with television set-top box **104** using any type of communication, such as a wired connection or any type of wireless communication (e.g., Bluetooth, Wi-Fi, radio frequency, infrared transmission, etc.), including via network **108**. In some examples, users can interact with television set-top box **104** through user device **102**, remote control **106**, or interface elements integrated within television set-top box **104** (e.g., buttons, a microphone, a camera, a joystick, etc.). For

example, speech input including media-related queries or commands for the virtual assistant can be received at user device **102** and/or remote control **106**, and the speech input can be used to cause media-related tasks to be executed on television set-top box **104**. Likewise, tactile commands for controlling media on television set-top box **104** can be received at user device **102** and/or remote control **106** (as well as from other devices not shown). The various functions of television set-top box **104** can thus be controlled in a variety of ways, giving users multiple options for controlling media content from multiple devices.

[0060] The client-side portion of the exemplary virtual assistant executed on user device **102** and/or television set-top box **104** with remote control **106** can provide client-side functionalities, such as user-facing input and output processing and communications with server system **110**. Server system **110** can provide server-side functionalities for any number of clients residing on a respective user device **102** or respective television set-top box **104**.

[0061] Server system **110** can include one or more virtual assistant servers **114** that can include a client-facing I/O interface **122**, one or more processing modules **118**, data and model storage **120**, and an I/O interface to external services **116**. The client-facing I/O interface **122** can facilitate the client-facing input and output processing for virtual assistant server **114**. The one or more processing modules **118** can utilize data and model storage **120** to determine the user's intent based on natural language input, and can perform task execution based on inferred user intent. In some examples, virtual assistant server **114** can communicate with external services **124**, such as telephony services, calendar services, information services, messaging services, navigation services, television programming services, streaming media services, and the like, through network(s) **108** for task completion or information acquisition. The I/O interface to external services **116** can facilitate such communications.

[0062] Server system **110** can be implemented on one or more standalone data processing devices or a distributed network of computers. In some examples, server system **110** can employ various virtual devices and/or services of third-party service providers (e.g., third-party cloud service providers) to provide the underlying computing resources and/or infrastructure resources of server system **110**.

[0063] Although the functionality of the virtual assistant is shown in FIG. **1** as including both a client-side portion and a server-side portion, in some examples, the functions of an assistant (or speech recognition and media control in general) can be implemented as a standalone application installed on a user device, television set-top box, smart television, or the like. In addition, the division of functionalities between the client and server portions of the virtual assistant can vary in different examples. For instance, in some examples, the client executed on user device **102** or television set-top box **104** can be a thin client that provides only user-facing input and output processing functions, and delegates all other functionalities of the virtual assistant to a backend server.

[0064] FIG. **2** illustrates a block diagram of exemplary user device **102** according to various examples. As shown, user device **102** can include a memory interface **202**, one or more processors **204**, and a peripherals interface **206**. The various components in user device **102** can be coupled together by one or more communication buses or signal lines. User device **102** can further include various sensors, subsystems, and peripheral devices that are coupled to the peripherals interface **206**. The sensors, subsystems, and peripheral devices can gather information and/or facilitate various functionalities of user device **102**.

[0065] For example, user device **102** can include a motion sensor **210**, a light sensor **212**, and a proximity sensor **214** coupled to peripherals interface **206** to facilitate orientation, light, and proximity sensing functions. One or more other sensors **216**, such as a positioning system (e.g., a GPS receiver), a temperature sensor, a biometric sensor, a gyroscope, a compass, an accelerometer, and the like, can also be connected to peripherals interface **206**, to facilitate related functionalities.

[0066] In some examples, a camera subsystem **220** and an optical sensor **222** can be utilized to facilitate camera functions, such as taking photographs and recording video clips. Communication functions can be facilitated through one or more wired and/or wireless communication subsystems **224**, which can include various communication ports, radio frequency receivers and transmitters, and/or optical (e.g., infrared) receivers and transmitters. An audio subsystem **226** can be coupled to speakers **228** and microphone **230** to facilitate voice-enabled functions, such as voice recognition, voice replication, digital recording, and telephony functions.

[0067] In some examples, user device **102** can further include an I/O subsystem **240** coupled to peripherals interface **206**. I/O subsystem **240** can include a touchscreen controller **242** and/or other input controller(s) **244**. Touchscreen controller **242** can be coupled to a touchscreen **246**. Touchscreen **246** and the touchscreen controller **242** can, for example, detect contact and movement or break thereof using any of a plurality of touch sensitivity technologies, such as capacitive, resistive, infrared, and surface acoustic wave technologies; proximity sensor arrays; and the like. Other input controller(s) **244** can be coupled to other input/control devices **248**, such as one or more buttons, rocker switches, a thumb-wheel, an infrared port, a USB port, and/or a pointer device, such as a stylus.

[0068] In some examples, user device **102** can further include a memory interface **202** coupled to memory **250**. Memory **250** can include any electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device; a portable computer diskette (magnetic); a random access memory (RAM) (magnetic); a read-only memory (ROM) (magnetic); an erasable programmable read-only memory (EPROM) (magnetic); a portable optical disc such as CD, CD-R, CD-RW, DVD, DVD-R, or DVD-RW; or flash memory such as compact flash cards, secured digital cards, USB memory devices, memory sticks, and the like. In some examples, a non-transitory computer-readable storage medium of memory **250** can be used to store instructions (e.g., for performing portions or all of the various processes described herein) for use by or in connection with an instruction execution system, apparatus, or device, such as a computer-based system, processor-containing system, or other system that can fetch the instructions from the instruction execution system, apparatus, or device, and can execute the instructions. In other examples, the instructions (e.g., for performing portions or all of the various processes described herein) can be stored on a non-transitory computer-readable storage medium of server system **110**, or can be divided between the non-transitory computer-readable storage medium of memory **250** and the non-transitory computer-readable storage medium of server system **110**. In the context of this document, a "non-transitory computer-readable stor-

age medium" can be any medium that can contain or store the program for use by or in connection with the instruction execution system, apparatus, or device.

[0069] In some examples, memory **250** can store an operating system **252**, a communication module **254**, a graphical user interface module **256**, a sensor processing module **258**, a phone module **260**, and applications **262**. Operating system **252** can include instructions for handling basic system services and for performing hardware-dependent tasks. Communication module **254** can facilitate communicating with one or more additional devices, one or more computers, and/or one or more servers. Graphical user interface module **256** can facilitate graphical user interface processing. Sensor processing module **258** can facilitate sensor-related processing and functions. Phone module **260** can facilitate phone-related processes and functions. Application module **262** can facilitate various functionalities of user applications, such as electronic messaging, web browsing, media processing, navigation, imaging, and/or other processes and functions.

[0070] As described herein, memory **250** can also store client-side virtual assistant instructions (e.g., in a virtual assistant client module **264**) and various user data **266** (e.g., user-specific vocabulary data, preference data, and/or other data such as the user's electronic address book, to-do lists, shopping lists, television program favorites, etc.) to, for example, provide the client-side functionalities of the virtual assistant. User data **266** can also be used in performing speech recognition in support of the virtual assistant or for any other application.

[0071] In various examples, virtual assistant client module **264** can be capable of accepting voice input (e.g., speech input), text input, touch input, and/or gestural input through various user interfaces (e.g., I/O subsystem **240**, audio subsystem **226**, or the like) of user device **102**. Virtual assistant client module **264** can also be capable of providing output in audio (e.g., speech output), visual, and/or tactile forms. For example, output can be provided as voice, sound, alerts, text messages, menus, graphics, videos, animations, vibrations, and/or combinations of two or more of the above. During operation, virtual assistant client module **264** can communicate with the virtual assistant server using communication subsystem **224**.

[0072] In some examples, virtual assistant client module **264** can utilize the various sensors, subsystems, and peripheral devices to gather additional information from the surrounding environment of user device **102** to establish a context associated with a user, the current user interaction, and/or the current user input. Such context can also include information from other devices, such as from television set-top box **104**. In some examples, virtual assistant client module **264** can provide the contextual information or a subset thereof with the user input to the virtual assistant server to help infer the user's intent. The virtual assistant can also use the contextual information to determine how to prepare and deliver outputs to the user. The contextual information can further be used by user device **102** or server system **110** to support accurate speech recognition.

[0073] In some examples, the contextual information that accompanies the user input can include sensor information, such as lighting, ambient noise, ambient temperature, images or videos of the surrounding environment, distance to another object, and the like. The contextual information can further include information associated with the physical state of user device **102** (e.g., device orientation, device location, device temperature, power level, speed, acceleration, motion patterns, cellular signal strength, etc.) or the software state of user device **102** (e.g., running processes, installed programs, past and present network activities, background services, error logs, resources usage, etc.). The contextual information can further include information associated with the state of connected devices or other devices associated with the user (e.g., media content displayed by television set-top box **104**, media content available to television set-top box **104**, etc.). Any of these types of contextual information can be provided to virtual assistant server **114** (or used on user device **102** itself) as contextual information associated with a user input.

[0074] In some examples, virtual assistant client module **264** can selectively provide information (e.g., user data **266**) stored on user device **102** in response to requests from virtual assistant server **114** (or it can be used on user device **102** itself in executing speech recognition and/or virtual assistant functions). Virtual assistant client module **264** can also elicit additional input from the user via a natural language dialogue or other user interfaces upon request by virtual assistant server **114**. Virtual assistant client module **264** can pass the additional input to virtual assistant server **114** to help virtual assistant server **114** in intent inference and/or fulfillment of the user's intent expressed in the user request.

[0075] In various examples, memory **250** can include additional instructions or fewer instructions. Furthermore, various functions of user device **102** can be implemented in hardware and/or in firmware, including in one or more signal processing and/or application specific integrated circuits.

[0076] FIG. **3** illustrates a block diagram of exemplary television set-top box **104** in system **300** for controlling television user interaction. System **300** can include a subset of the elements of system **100**. In some examples, system **300** can execute certain functions alone and can function together with other elements of system **100** to execute other functions. For example, the elements of system **300** can process certain media control functions without interacting with server system **110** (e.g., playback of locally stored media, recording functions, channel tuning, etc.), and system **300** can process other media control functions in conjunction with server system **110** and other elements of system **100** (e.g., playback of remotely stored media, downloading media content, processing certain virtual assistant queries, etc.). In other examples, the elements of system **300** can perform the functions of the larger system **100**, including accessing external services **124** through a network. It should be understood that functions can be divided between local devices and remote server devices in a variety of other ways.

[0077] As shown in FIG. **3**, in one example, television set-top box **104** can include memory interface **302**, one or more processors **304**, and a peripherals interface **306**. The various components in television set-top box **104** can be coupled together by one or more communication buses or signal lines. Television set-top box **104** can further include various subsystems and peripheral devices that are coupled to the peripherals interface **306**. The subsystems and peripheral devices can gather information and/or facilitate various functionalities of television set-top box **104**.

[0078] For example, television set-top box **104** can include a communications subsystem **324**. Communication functions can be facilitated through one or more wired and/or wireless communication subsystems **324**, which can include various

communication ports, radio frequency receivers and transmitters, and/or optical (e.g., infrared) receivers and transmitters.

[0079] In some examples, television set-top box **104** can further include an I/O subsystem **340** coupled to peripherals interface **306**. I/O subsystem **340** can include an audio/video output controller **370**. Audio/video output controller **370** can be coupled to a display **112** and speakers **111** or can otherwise provide audio and video output (e.g., via audio/video ports, wireless transmission, etc.). I/O subsystem **340** can further include remote controller **342**. Remote controller **342** can be communicatively coupled to remote control **106** (e.g., via a wired connection, Bluetooth, Wi-Fi, etc.). Remote control **106** can include microphone **372** for capturing audio input (e.g., speech input from a user), button(s) **374** for capturing tactile input, and transceiver **376** for facilitating communication with television set-top box **104** via remote controller **342**. Remote control **106** can also include other input mechanisms, such as a keyboard, joystick, touchpad, or the like. Remote control **106** can further include output mechanisms, such as lights, a display, a speaker, or the like. Input received at remote control **106** (e.g., user speech, button presses, etc.) can be communicated to television set-top box **104** via remote controller **342**. I/O subsystem **340** can also include other input controller(s) **344**. Other input controller(s) **344** can be coupled to other input/control devices **348**, such as one or more buttons, rocker switches, a thumb-wheel, an infrared port, a USB port, and/or a pointer device, such as a stylus.

[0080] In some examples, television set-top box **104** can further include a memory interface **302** coupled to memory **350**. Memory **350** can include any electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device; a portable computer diskette (magnetic); a random access memory (RAM) (magnetic); a read-only memory (ROM) (magnetic); an erasable programmable read-only memory (EPROM) (magnetic); a portable optical disc such as CD, CD-R, CD-RW, DVD, DVD-R, or DVD-RW; or flash memory such as compact flash cards, secured digital cards, USB memory devices, memory sticks, and the like. In some examples, a non-transitory computer-readable storage medium of memory **350** can be used to store instructions (e.g., for performing portions or all of the various processes described herein) for use by or in connection with an instruction execution system, apparatus, or device, such as a computer-based system, processor-containing system, or other system that can fetch the instructions from the instruction execution system, apparatus, or device, and can execute the instructions. In other examples, the instructions (e.g., for performing portions or all of the various processes described herein) can be stored on a non-transitory computer-readable storage medium of server system **110**, or can be divided between the non-transitory computer-readable storage medium of memory **350** and the non-transitory computer-readable storage medium of server system **110**. In the context of this document, a "non-transitory computer-readable storage medium" can be any medium that can contain or store the program for use by or in connection with the instruction execution system, apparatus, or device.

[0081] In some examples, memory **350** can store an operating system **352**, a communication module **354**, a graphical user interface module **356**, an on-device media module **358**, an off-device media module **360**, and applications **362**. Operating system **352** can include instructions for handling basic system services and for performing hardware-dependent tasks. Communication module **354** can facilitate communicating with one or more additional devices, one or more computers, and/or one or more servers. Graphical user interface module **356** can facilitate graphical user interface processing. On-device media module **358** can facilitate storage and playback of media content stored locally on television set-top box **104** and other media content available locally (e.g., cable channel tuning). Off-device media module **360** can facilitate streaming playback or download of media content stored remotely (e.g., on a remote server, on user device **102**, etc.). Application module **362** can facilitate various functionalities of user applications, such as electronic messaging, web browsing, media processing, gaming, and/or other processes and functions.

[0082] As described herein, memory **350** can also store client-side virtual assistant instructions (e.g., in a virtual assistant client module **364**) and various user data **366** (e.g., user-specific vocabulary data, preference data, and/or other data such as the user's electronic address book, to-do lists, shopping lists, television program favorites, etc.) to, for example, provide the client-side functionalities of the virtual assistant. User data **366** can also be used in performing speech recognition in support of the virtual assistant or for any other application.

[0083] In various examples, virtual assistant client module **364** can be capable of accepting voice input (e.g., speech input), text input, touch input, and/or gestural input through various user interfaces (e.g., I/O subsystem **340** or the like) of television set-top box **104**. Virtual assistant client module **364** can also be capable of providing output in audio (e.g., speech output), visual, and/or tactile forms. For example, output can be provided as voice, sound, alerts, text messages, menus, graphics, videos, animations, vibrations, and/or combinations of two or more of the above. During operation, virtual assistant client module **364** can communicate with the virtual assistant server using communication subsystem **324**.

[0084] In some examples, virtual assistant client module **364** can utilize the various subsystems and peripheral devices to gather additional information from the surrounding environment of television set-top box **104** to establish a context associated with a user, the current user interaction, and/or the current user input. Such context can also include information from other devices, such as from user device **102**. In some examples, virtual assistant client module **364** can provide the contextual information or a subset thereof with the user input to the virtual assistant server to help infer the user's intent. The virtual assistant can also use the contextual information to determine how to prepare and deliver outputs to the user. The contextual information can further be used by television set-top box **104** or server system **110** to support accurate speech recognition.

[0085] In some examples, the contextual information that accompanies the user input can include sensor information, such as lighting, ambient noise, ambient temperature, distance to another object, and the like. The contextual information can further include information associated with the physical state of television set-top box **104** (e.g., device location, device temperature, power level, etc.) or the software state of television set-top box **104** (e.g., running processes, installed applications, past and present network activities, background services, error logs, resources usage, etc.). The contextual information can further include information associated with the state of connected devices or other devices associated with the user (e.g., content displayed on user

device **102**, playable content on user device **102**, etc.). Any of these types of contextual information can be provided to virtual assistant server **114** (or used on television set-top box **104** itself) as contextual information associated with a user input.

[0086] In some examples, virtual assistant client module **364** can selectively provide information (e.g., user data **366**) stored on television set-top box **104** in response to requests from virtual assistant server **114** (or it can be used on television set-top box **104** itself in executing speech recognition and/or virtual assistant functions). Virtual assistant client module **364** can also elicit additional input from the user via a natural language dialogue or other user interfaces upon request by virtual assistant server **114**. Virtual assistant client module **364** can pass the additional input to virtual assistant server **114** to help virtual assistant server **114** in intent inference and/or fulfillment of the user's intent expressed in the user request.

[0087] In various examples, memory **350** can include additional instructions or fewer instructions. Furthermore, various functions of television set-top box **104** can be implemented in hardware and/or in firmware, including in one or more signal processing and/or application specific integrated circuits.

[0088] It should be understood that system **100** and system **300** are not limited to the components and configuration shown in FIG. **1** and FIG. **3**, and user device **102**, television set-top box **104**, and remote control **106** are likewise not limited to the components and configuration shown in FIG. **2** and FIG. **3**. System **100**, system **300**, user device **102**, television set-top box **104**, and remote control **106** can all include fewer or other components in multiple configurations according to various examples.

[0089] Throughout this disclosure, references to "the system" can include system **100**, system **300**, or one or more elements of either system **100** or system **300**. For example, a typical system referred to herein can include at least television set-top box **104** receiving user input from remote control **106** and/or user device **102**.

[0090] FIGS. **4**A through **4**E illustrate exemplary speech input interface **484** that can be shown on a display (such as display **112**) to convey speech input information to a user. In one example, speech input interface **484** can be shown over video **480**, which can include any moving images or paused video. For example, video **480** can include live television, a playing video, a streaming movie, playback of a recorded program, or the like. Speech input interface **484** can be configured to occupy a minimal amount of space so as not to significantly interfere with user viewing of video **480**.

[0091] In one example, a virtual assistant can be triggered to listen for speech input containing a command or query (or to commence recording of speech input for subsequent processing or commence processing in real-time of speech input). Listening can be triggered in a variety of ways, including indications such as a user pressing a physical button on remote control **106**, a user pressing a physical button on user device **102**, a user pressing a virtual button on user device **102**, a user uttering a trigger phrase that is recognizable by an always-listening device (e.g., uttering "Hey Assistant" to commence listening for a command), a user performing a gesture detectable by a sensor (e.g., motioning in front of a camera), or the like. In another example, a user can press and hold a physical button on remote control **106** or user device **102** to initiate listening. In still other examples, a user can press and hold a physical button on remote control **106** or user device **102** while speaking a query or command, and can release the button when finished. Various other indications can likewise be received to initiate receipt of speech input from the user.

[0092] In response to receiving an indication to listen for speech input, speech input interface **484** can be displayed. FIG. **4**A illustrates notification area **482** expanding upward from the bottom portion of display **112**. Speech input interface **484** can be displayed in notification area **482** upon receipt of an indication to listen for speech input, and the interface can be animated to slide upward from the bottom edge of the viewing area of display **112** as shown. FIG. **4**B illustrates speech input interface **484** after sliding upward into view. Speech input interface **484** can be configured to occupy a minimal amount of space at the bottom of display **112** to avoid significantly interfering with video **480**. In response to receiving the indication to listen for speech input, readiness confirmation **486** can be displayed. Readiness confirmation **486** can include a microphone symbol as shown, or can include any other image, icon, animation, or symbol to convey that the system (e.g., one or more elements of system **100**) is ready to capture speech input from the user.

[0093] As the user begins to speak, listening confirmation **487** shown in FIG. **4**C can be displayed to confirm that the system is capturing the speech input. In some examples, listening confirmation **487** can be displayed in response to receiving speech input (e.g., capturing speech). In other examples, readiness confirmation **486** can be displayed for a predetermined amount of time (e.g., 500 milliseconds, 1 second, 3 seconds, etc.) after which listening confirmation **487** can be displayed. Listening confirmation **487** can include a waveform symbol as shown, or can include an active waveform animation that moves (e.g., changes frequency) in response to user speech. In other examples, listening confirmation **487** can include any other image, icon, animation, or symbol to convey that the system is capturing speech input from the user.

[0094] Upon detecting that the user has finished speaking (e.g., based on a pause, speech interpretation indicating the end of a query, or any other endpoint detection method), processing confirmation **488** shown in FIG. **4**D can be displayed to confirm that the system finished capturing the speech input and is processing the speech input (e.g., interpreting the speech input, determining user intent, and/or executing the associated tasks). Processing confirmation **488** can include an hourglass symbol as shown, or can include any other image, icon, animation, or symbol to convey that the system is processing the captured speech input. In another example, processing confirmation **488** can include an animation of a spinning circle or a colored/glowing point moving around a circle.

[0095] After the captured speech input is interpreted as text (or in response to successfully converting the speech input to text), command receipt confirmation **490** and/or transcription **492** shown in FIG. **4**E can be displayed to confirm that the system received and interpreted the speech input. Transcription **492** can include a transcription of the received speech input (e.g., "What sporting events are on right now?"). In some examples, transcription **492** can be animated to slide up from the bottom of display **112**, can be displayed momentarily in the position shown in FIG. **4**E (e.g., a few seconds), and can then be slid up to the top of speech input interface **484** before disappearing from view (e.g., as though the text is

scrolled up and eventually out of view). In other examples, a transcription may not be displayed, and the user's command or query can be processed and associated tasks can be executed without displaying a transcription (e.g., a simple channel change can be executed immediately without displaying a transcription of the user's speech).

[0096] In other examples, speech transcription can be performed in real-time as a user speaks. As words are transcribed, they can be displayed in speech input interface **484**. For example, the words can be displayed alongside listening confirmation **487**. After the user finishes speaking, command receipt confirmation **490** can be displayed briefly before executing the tasks associated with the user's command.

[0097] Moreover, in other examples, command receipt confirmation **490** can convey information about received and understood commands. For example, for a simple request to change to another channel, a logo or number associated with the channel can briefly be displayed as command receipt confirmation **490** (e.g., for a few seconds) as the channel is changed. In another example, for a request to pause a video (e.g., video **480**), a pause symbol (e.g., two vertical, parallel bars) can be displayed as command receipt confirmation **490**. The pause symbol can remain on the display until, for example, the user performs another action (e.g., issuing a play command to resume playback). Symbols, logos, animations, or the like can likewise be displayed for any other command (e.g., symbols for rewind, fast forward, stop, play, etc.). Command receipt confirmation **490** can thus be used to convey command-specific information.

[0098] In some examples, speech input interface **484** can be hidden after receipt of a user query or command. For example, speech input interface **484** can be animated as sliding downward until it is out of view of the bottom of display **112**. Speech input interface **484** can be hidden in instances where further information need not be displayed to the user. For example, for common or straightforward commands (e.g., change to channel ten, change to the sports channel, play, pause, fast forward, rewind, etc.), speech input interface **484** can be hidden immediately after confirming command receipt, and the associated task or tasks can be performed immediately. Although various examples herein illustrate and describe an interface at a bottom or top edge of a display, it should be appreciated that any of the various interfaces can be positioned in other locations around a display. For example, speech input interface **484** can emerge from a side edge of display **112**, in the center of display **112**, in a corner of display **112**, or the like. Similarly, the various other interface examples described herein can be arranged in a variety of different orientations in a variety of different locations on a display. Moreover, although various interfaces described herein are illustrated as opaque, any of the various interfaces can be transparent or otherwise allow an image (blurred or whole) to be viewed through the interface (e.g., overlaying interface content on media content without completely obscuring the underlying media content).

[0099] In other examples, the result of a query can be displayed within speech input interface **484** or in a different interface. FIG. **5** illustrates exemplary media content interface **510** over video **480** with an exemplary result of the transcribed query of FIG. **4**E. In some examples, the result of a virtual assistant query can include media content instead of or in addition to textual content. For example, the result of a virtual assistant query can include television programs, videos, music, or the like. Some results can include media immediately available for playback, while other results can include media that may be available for purchase or the like.

[0100] As shown, media content interface **510** can be a larger size than speech input interface **484**. In one example, speech input interface **484** can be of a smaller first size to accommodate speech input information, while media content interface **510** can be of a larger second size to accommodate query results, which can include text, still images, and moving images. In this manner, interfaces for conveying virtual assistant information can scale in size according to the content that is to be conveyed, thereby limiting screen real estate intrusion (e.g., minimally blocking other content, such as video **480**).

[0101] As illustrated, media content interface **510** can include (as a result of a virtual assistant query) selectable video links **512**, selectable text links **514**, and additional content link **513**. In some examples, links can be selected by navigating focus, a cursor, or the like to a particular element and selecting it using a remote control (e.g., remote control **106**). In other examples, links can be selected using voice commands to the virtual assistant (e.g., watch that soccer game, show details about the basketball game, etc.). Selectable video links **512** can include still or moving images and can be selectable to cause playback of the associated video. In one example, selectable video link **512** can include a playing video of the associated video content. In another example, selectable video link **512** can include a live feed of a television channel. For example, selectable video link **512** can include a live feed of a soccer game on a sports channel as a result of a virtual assistant query about sporting events currently on television. Selectable video link **512** can also include any other video, animation, image, or the like (e.g., a triangular play symbol). Moreover, link **512** can link to any type of media content, such as a movie, television show, sporting event, music, or the like.

[0102] Selectable text links **514** can include textual content associated with selectable video links **512** or can include textual representations of results of a virtual assistant query. In one example, selectable text links **514** can include a description of media resulting from a virtual assistant query. For instance, selectable text link **514** can include the name of a television program, title of a movie, description of a sporting event, television channel name or number, or the like. In one example, selection of text link **514** can cause playback of the associated media content. In another example, selection of text link **514** can provide additional detailed information about the media content or other virtual assistant query result. Additional content link **513** can link to and cause display of additional results of a virtual assistant query.

[0103] Although certain media content examples are shown in FIG. **5**, it should be appreciated that any type of media content can be included as a result of a virtual assistant query for media content. For example, media content that can be returned as a result of a virtual assistant can include videos, television programs, music, television channels, or the like. In addition, in some examples, categorical filters can be provided in any of the interfaces herein to allow users to filter search or query results or displayed media options. For example, selectable filters can be provided to filter results by type (e.g., movies, music albums, books, television shows, etc.). In other examples, selectable filters can include genre or content descriptors (e.g., comedy, interview, specific program, etc.). In still other examples, selectable filters can include times (e.g., this week, last week, last year, etc.). It

should be appreciated that filters can be provided in any of the various interfaces described herein to allow users to filter results based on categories relevant to the displayed content (e.g., filter by type where media results have various types, filter by genre where media results have various genres, filter by times where media results have various times, etc.).

[0104] In other examples, media content interface **510** can include a paraphrase of a query in addition to media content results. For example, a paraphrase of the user's query can be displayed above the media content results (above selectable video links **512** and selectable text links **514**). In the example of FIG. **5**, such a paraphrase of the user's query can include the following: "Here are some sporting events that are on right now." Other text introducing the media content results can likewise be displayed.

[0105] In some examples, after displaying any interface, including interface **510**, a user can initiate capture of additional speech input with a new query (that may or may not be related to previous queries). User queries can include commands to act on interface elements, such as a command to select a video link **512**. In another example, user speech can include a query associated with displayed content, such as displayed menu information, a playing video (e.g., video **480**), or the like. A response can be determined for such a query based on the information shown (e.g., displayed text) and/or metadata associated with displayed content (e.g., metadata associated with a playing video). For example, a user can ask about a media result shown in an interface (e.g., interface **510**), and metadata associated with that media can be searched to provide an answer or result. Such an answer or result can then be provided in another interface or within the same interface (e.g., in any of the interfaces discussed herein).

[0106] As noted above, in one example, additional detailed information about media content can be displayed in response to selection of a text link **514**. FIGS. **6**A and **6**B illustrate exemplary media detail interface **618** over video **480** after selection of a text link **514**. In one example, in providing addition detailed information, media content interface **510** can be expanded into media detail interface **618** as illustrated by interface expansion transition **616** of FIG. **6**A. In particular, as shown in FIG. **6**A, selected content can be expanded in size and additional textual information can be provided by expanding the interface upward on display **112** to occupy more of the screen real estate. The interface can be expanded to accommodate the additional detailed information desired by the user. In this manner, the size of the interface can scale with the amount of content desired by the user, thereby minimizing screen real estate intrusion while still conveying the desired content.

[0107] FIG. **6**B illustrates detail interface **618** after full expansion. As shown, detail interface **618** can be of a larger size than either media content interface **510** or speech input interface **484** to accommodate the desired detailed information. Detail interface **618** can include detailed media information **622**, which can include a variety of detailed information associated with media content or another result of a virtual assistant query. Detailed media information **622** can include a program title, program description, program air time, channel, episode synopsis, movie description, actor names, character names, sporting event participants, producer names, director names, or any other detailed information associated with a result of a virtual assistant query.

[0108] In one example, detail interface **618** can include selectable video link **620** (or another link to play media content), which can include a larger version of a corresponding selectable video link **512**. As such, selectable video link **620** can include still or moving images and can be selectable to cause playback of the associated video. Selectable video link **620** can include a playing video of the associated video content, a live feed of a television channel (e.g., a live feed of a soccer game on a sports channel), or the like. Selectable video link **620** can also include any other video, animation, image, or the like (e.g., a triangular play symbol).

[0109] As noted above, a video can be played in response to selection of a video link, such as video link **620** or video links **512**. FIGS. **7**A and **7**B illustrate an exemplary media transition interface that can be displayed in response to selection of a video link (or other command to play video content). As illustrated, video **480** can be replaced with video **726**. In one example, video **726** can be expanded to overtake or cover video **480** as shown by interface expansion transition **724** in FIG. **7**A. The result of the transition can include expanded media interface **728** of FIG. **7**B. As with other interfaces, the size of expanded media interface **728** can be sufficient to provide the user with the desired information; here, that can include expanding to fill display **112**. Expanded media interface **728** can thus be larger than any other interface as the desired information can include playing media content across the entire display. Although not shown, in some examples, descriptive information can briefly be overlaid on video **726** (e.g., along the bottom of the screen). Such descriptive information can include the name of the associated program, video, channel, or the like. The descriptive information can then be hidden from view (e.g., after a few seconds).

[0110] FIGS. **8**A and **8**B illustrate exemplary speech input interface **836** that can be shown on display **112** to convey speech input information to a user. In one example, speech input interface **836** can be shown over menu **830**. Menu **830** can include various media options **832**, and speech input interface **836** can similarly be displayed over any other type of menu (e.g., content menus, category menus, control menus, setup menus, program menus, etc.). In one example, speech input interface **836** can be configured to occupy a relatively large amount of screen real estate of display **112**. For example, speech input interface **836** can be larger than speech input interface **484** discussed above. In one example, the size of speech input interface to use (e.g., either the smaller interface **484** or the larger interface **836**) can be determined based on the background content. When the background content includes a moving image, for example, a small size speech input interface can be displayed (e.g., interface **484**). On the other hand, when the background content includes a still image (e.g., a paused video) or a menu, for example, a large size speech input interface can be displayed (e.g., interface **836**). In this manner, if a user is watching video content, a smaller speech input interface can be displayed that only minimally intrudes on the screen real estate; whereas if a user is navigating a menu or viewing a paused video or other still image, a larger speech input interface can be displayed that can convey more information or have a more profound effect by occupying additional real estate. Other interfaces discussed herein can likewise be sized differently based on background content.

[0111] As discussed above, a virtual assistant can be triggered to listen for speech input containing a command or query (or to commence recording of speech input for subsequent processing or commence processing in real-time of speech input). Listening can be triggered in a variety of ways,

including indications such as a user pressing a physical button on remote control **106**, a user pressing a physical button on user device **102**, a user pressing a virtual button on user device **102**, a user uttering a trigger phrase that is recognizable by an always-listening device (e.g., uttering "Hey Assistant" to commence listening for a command), a user performing a gesture detectable by a sensor (e.g., motioning in front of a camera), or the like. In another example, a user can press and hold a physical button on remote control **106** or user device **102** to initiate listening. In still other examples, a user can press and hold a physical button on remote control **106** or user device **102** while speaking a query or command, and can release the button when finished. Various other indications can likewise be received to initiate receipt of speech input from the user.

[0112] In response to receiving an indication to listen for speech input, speech input interface **836** can be displayed over menu **830**. FIG. **8**A illustrates large notification area **834** expanding upward from the bottom portion of display **112**. Speech input interface **836** can be displayed in large notification area **834** upon receipt of an indication to listen for speech input, and the interface can be animated to slide upward from the bottom edge of the viewing area of display **112** as shown. In some examples, a background menu, paused video, still image, or other background content can be contracted and/or moved backward in the z direction (as if further into display **112**) as an overlapping interface is displayed (e.g., in response to receiving an indication to listen for speech input). Background interface contraction transition **831** and the associated inward-pointing arrows illustrate how background content (e.g., menu **830**) can be contracted—shrinking the displayed menu, images, text, etc. This can provide a visual effect of the background content appearing to move away from the user, out of the way of a new foreground interface (e.g., interface **836**). FIG. **8**B illustrates contracted background interface **833**, including a contracted (shrunken) version of menu **830**. As shown, contracted background interface **833** (which can include a border) can appear further from the user while ceding focus to the foreground interface **836**. Background content in any of the other examples discussed herein (including background video content) can similarly be contracted and/or moved backward in the z direction as overlapping interfaces are displayed.

[0113] FIG. **8**B illustrates speech input interface **836** after sliding upward into view. As discussed above, various confirmations can be displayed while receiving speech input. Although not shown here, speech input interface **836** can similarly display larger versions of readiness confirmation **486**, listening confirmation **487**, and/or processing confirmation **488** in a similar manner as speech input interface **484** discussed above with reference to FIGS. **4**B, **4**C, and **4**D, respectively.

[0114] As shown in FIG. **8**B, command receipt confirmation **838** can be shown (as with the smaller sized command receipt confirmation **490** discussed above) to confirm that the system received and interpreted the speech input. Transcription **840** can also be shown and can include a transcription of the received speech input (e.g., "What's the weather in New York?"). In some examples, transcription **840** can be animated to slide up from the bottom of display **112**, can be displayed momentarily in the position shown in FIG. **8**B (e.g., a few seconds), and can then be slid up to the top of speech input interface **836** before disappearing from view (e.g., as though the text is scrolled up and eventually out of view). In other examples, a transcription may not be displayed, and the user's command or query can be processed and associated tasks can be executed without displaying a transcription.

[0115] In other examples, speech transcription can be performed in real-time as a user speaks. As words are transcribed, they can be displayed in speech input interface **836**. For example, the words can be displayed alongside a larger version of listening confirmation **487** discussed above. After the user finishes speaking, command receipt confirmation **838** can be displayed briefly before executing the tasks associated with the user's command.

[0116] Moreover, in other examples, command receipt confirmation **838** can convey information about received and understood commands. For example, for a simple request to tune to a particular channel, a logo or number associated with the channel can briefly be displayed as command receipt confirmation **838** (e.g., for a few seconds) as the channel is tuned. In another example, for a request to select a displayed menu item (e.g., one of media options **832**), an image associated with the selected menu item can be displayed as command receipt confirmation **838**. Command receipt confirmation **838** can thus be used to convey command-specific information.

[0117] In some examples, speech input interface **836** can be hidden after receipt of a user query or command. For example, speech input interface **836** can be animated as sliding downward until it is out of view of the bottom of display **112**. Speech input interface **836** can be hidden in instances where further information need not be displayed to the user. For example, for common or straightforward commands (e.g., change to channel ten, change to the sports channel, play that movie, etc.), speech input interface **836** can be hidden immediately after confirming command receipt, and the associated task or tasks can be performed immediately.

[0118] In other examples, the result of a query can be displayed within speech input interface **836** or in a different interface. FIG. **9** illustrates exemplary virtual assistant result interface **942** over menu **830** (specifically over contracted background interface **833**) with an exemplary result of the transcribed query of FIG. **8**B. In some examples, the result of a virtual assistant query can include a textual answer, such as text answer **944**. The result of a virtual assistant query can also include media content that addresses a user's query, such as the content associated with selectable video link **946** and purchase link **948**. In particular, in this example, a user can ask for weather information for the specified location of New York. The virtual assistant can provide text answer **944** directly answering the user's query (e.g., indicating that the weather looks good and providing temperature information). Instead of or in addition to text answer **944**, the virtual assistant can provide selectable video link **946** along with purchase link **948** and the associated text. The media associated with links **946** and **948** can also provide a response to the user's query. Here, the media associated with links **946** and **948** can include a ten-minute clip of weather information at the specified location—specifically, the five-day forecast for New York from a television channel called the Weather Forecast Channel.

[0119] In one example, the clip addressing the user's query can include a time-cued portion of previously-aired content (that may be available from a recording or from a streaming service). The virtual assistant can, in one example, identify such content based on the user intent associated with the speech input and by searching detailed information about

available media content (e.g., including metadata for recorded programs along with detailed timing information or detailed information about streaming content). In some examples, a user may not have access to or may not have a subscription for certain content. In such instances, content can be offered for purchase, such as via purchase link **948**. The cost of the content can be automatically withdrawn from a user account or charged to a user account upon selection of purchase link **948** or video link **946**.

[0120] FIG. **10** illustrates exemplary process **1000** for controlling television interactions using a virtual assistant and displaying associated information using different interfaces. At block **1002**, speech input can be received from a user. For example, speech input can be received at user device **102** or remote control **106** of system **100**. In some examples, the speech input (or a data representation of some or all of the speech input) can be transmitted to and received by server system **110** and/or television set-top box **104**. In response to a user initiating receipt of speech input, various notifications can be displayed on a display (such as display **112**). For example, a readiness confirmation, listening confirmation, processing confirmation, and/or command receipt confirmation can be displayed as discussed above with reference to FIGS. **4**A-**4**E. In addition, received user speech input can be transcribed, and a transcription can be displayed.

[0121] Referring again to process **1000** of FIG. **10**, at block **1004**, media content can be determined based on the speech input. For example, media content that addresses a user query directed at a virtual assistant can be determined (e.g., by searching available media content or the like). For instance, media content can be determined related to transcription **492** of FIG. **4**E ("What sporting events are on right now?"). Such media content can include live sporting events being shown on one or more television channels available to the user for viewing.

[0122] At block **1006**, a first user interface of a first size with selectable media links can be displayed. For example, media content interface **510** with selectable video links **512** and selectable text links **514** can be displayed on display **112** as shown in FIG. **5**. As discussed above, media content interface **510** can be of a smaller size to avoid interfering with background video content.

[0123] At block **1008**, a selection of one of the links can be received. For example, selection of one of links **512** and/or links **514** can be received. At block **1010**, a second user interface of a larger second size with media content associated with the selection can be displayed. For example, detail interface **618** with selectable video link **620** and detailed media information **622** can be displayed as shown in FIG. **6**B. As discussed above, detail interface **618** can be of a larger size to convey the desired additional detailed media information. Similarly, upon selection of video link **620**, expanded media interface **728** can be displayed with video **726** as shown in FIG. **7**B. As discussed above, expanded media interface **728** can be of a larger size still to provide the desired media content to the user. In this manner, the various interfaces discussed herein can be sized to accommodate desired content (including expanding into larger sized interfaces or contracting down to smaller sized interfaces) while otherwise occupying limited screen real estate. Process **1000** can thus be used to control television interactions using a virtual assistant and display associated information using different interfaces.

[0124] In another example, a larger size interface can be displayed over a control menu than over background video content. For example, speech input interface **836** can be displayed over menu **830** as shown in FIG. **8**B, and assistant result interface **942** can be displayed over menu **830** as shown in FIG. **9**, whereas smaller media content interface **510** can be displayed over video **480** as shown in FIG. **5**. In this manner, the size of an interface (e.g., the amount of screen real estate occupied by an interface) can be determined, at least in part, by the type of background content.

[0125] FIG. **11** illustrates exemplary television media content on user device **102**, which can include a mobile telephone, tablet computer, remote control, or the like with touchscreen **246** (or another display). FIG. **11** illustrates interface **1150** including a TV listing with multiple television programs **1152**. Interface **1150** can, for example, correspond to a particular application on user device **102**, such as a television control application, television content listing application, Internet application, or the like. In some examples, content shown on user device **102** (e.g., on touchscreen **246**) can be used to determine user intent from speech input relating to that content, and the user intent can be used to cause playback or display of content on another device and display (e.g., on television set-top box **104** and display **112** and/or speakers **111**). For example, content shown in interface **1150** on user device **102** can be used to disambiguate user requests and determine user intent from speech input, and the determined user intent can then be used to play or display media via television set-top box **104**.

[0126] FIG. **12** illustrates exemplary television control using a virtual assistant. FIG. **12** illustrates interface **1254**, which can include a virtual assistant interface formatted as a conversational dialog between the assistant and the user. For example, interface **1254** can include assistant greeting **1256** prompting the user to make a request. Subsequently-received user speech can then be transcribed, such as transcribed user speech **1258**, showing the back and forth conversation. In some examples, interface **1254** can appear on user device **102** in response to a trigger to initiate receipt of speech input (triggers such as button presses, key phrases, or the like).

[0127] In one example, a user request to play content via television set-top box **104** (e.g., on display **112** and speakers **111**) can include an ambiguous reference to something shown on user device **102**. Transcribed user speech **1258**, for example, includes a reference to "that" soccer game ("Put on that soccer game."). The particular soccer game desired can be unclear from the speech input alone. In some examples, however, the content shown on user device **102** can be used to disambiguate user requests and determine user intent. In one example, content shown on user device **102** prior to the user making the request (e.g., prior to interface **1254** appearing on touchscreen **246**) can be used to determine user intent (as can content appearing within interface **1254**, such as previous queries and results). In the illustrated example, the content shown in interface **1150** of FIG. **11** can be used to determine the user intent from the command to put on "that" soccer game. The TV listing of television programs **1152** includes a variety of different programs, one of which is titled "Soccer" appearing on Channel 5. The appearance of the soccer listing can be used to determine the user's intent from uttering "that" soccer game. In particular, the user's reference to "that" soccer game can be resolved to the soccer program appearing in the TV listing of interface **1150**. Accordingly, the virtual assistant can cause playback of that particular soccer game that the user desired (e.g., by causing television set-top box **104** to tune to the appropriate channel and show the game).

[0128] In other examples, a user can reference television programs shown in interface **1150** in a variety of other ways (e.g., the show on channel eight, the news, the drama show, the advertisement, the first show, etc.), and user intent can similarly be determined based on displayed content. It should be appreciated that metadata associated with displayed content (e.g., TV program descriptions), fuzzy matching techniques, synonym matching, and the like can further be used in conjunction with displayed content to determine user intent. For example, the term "advertisement" can be matched to the description "paid programming" (e.g., using synonyms and/or fuzzy matching techniques) to determine user intent from a request to show "the advertisement." Likewise, the description of a particular TV program can be analyzed in determining user intent. For example, the term "law" could be identified in the detailed description of a courtroom drama, and the user intent can be determined from a user request to watch the "law" show based on the detailed description associated with the content shown in interface **1150**. Displayed content and data associated with it can thus be used to disambiguate user requests and determine user intent.

[0129] FIG. **13** illustrates exemplary picture and video content on user device **102**, which can include a mobile telephone, tablet computer, remote control, or the like with touchscreen **246** (or another display). FIG. **13** illustrates interface **1360** including a listing of photos and videos. Interface **1360** can, for example, correspond to a particular application on user device **102**, such as a media content application, file navigation application, storage application, remote storage management application, camera application, or the like. As shown, interface **1360** can include video **1362**, photo album **1364** (e.g., a group of multiple photos), and photos **1366**. As discussed above with reference to FIG. **11** and FIG. **12**, content shown on user device **102** can be used to determine user intent from speech input relating to that content. The user intent can then be used to cause playback or display of content on another device and display (e.g., on television set-top box **104** and display **112** and/or speakers **111**). For example, content shown in interface **1360** on user device **102** can be used to disambiguate user requests and to determine user intent from speech input, and the determined user intent can then be used to play or display media via television set-top box **104**.

[0130] FIG. **14** illustrates exemplary media display control using a virtual assistant. FIG. **14** illustrates interface **1254**, which can include a virtual assistant interface formatted as a conversational dialog between the assistant and the user. As shown, interface **1254** can include assistant greeting **1256** prompting the user to make a request. Within the dialog, user speech can then be transcribed as shown by the examples of FIG. **14**. In some examples, interface **1254** can appear on user device **102** in response to a trigger to initiate receipt of speech input (triggers such as button presses, key phrases, or the like).

[0131] In one example, a user request to play media content or display media via television set-top box **104** (e.g., on display **112** and speakers **111**) can include an ambiguous reference to something shown on user device **102**. Transcribed user speech **1468**, for example, includes a reference to "that" video ("Show that video."). The particular video referenced can be unclear from the speech input alone. In some examples, however, the content shown on user device **102** can be used to disambiguate user requests and determine user intent. In one example, content shown on user device **120** prior to the user making the request (e.g., prior to interface **1254** appearing on touchscreen **246**) can be used to determine user intent (as can content appearing within interface **1254**, such as previous queries and results). In the example of user speech **1468**, the content shown in interface **1360** of FIG. **13** can be used to determine the user intent from the command to show "that" video. The listing of photos and videos in interface **1360** includes a variety of different photos and a video, including video **1362**, photo album **1354**, and photos **1366**. As only one video appears in interface **1360** (e.g., video **1362**), the appearance of video **1362** in interface **1360** can be used to determine the user's intent from uttering "that" video. In particular, the user's reference to "that" video can be resolved to video **1362** (titled "Graduation Video") appearing in interface **1360**. Accordingly, the virtual assistant can cause playback of video **1362** (e.g., by causing video **1362** to be transmitted to television set-top box **104** from user device **102** or remote storage and causing playback to commence).

[0132] In another example, transcribed user speech **1470** includes a reference to "that" album ("Play a slideshow of that album."). The particular album referenced can be unclear from the speech input alone. The content shown on user device **102** can again be used to disambiguate the user request. In particular, the content shown in interface **1360** of FIG. **13** can be used to determine the user intent from the command to play a slideshow of "that" album. The listing of photos and videos in interface **1360** includes photo album **1354**. The appearance of photo album **1364** in interface **1360** can be used to determine the user's intent from uttering "that" album. In particular, the user's reference to "that" album can be resolved to photo album **1364** (titled "Graduation Album") appearing in interface **1360**. In response to user speech **1470**, therefore, the virtual assistant can cause a slideshow to be displayed including the photos from photo album **1364** (e.g., by causing the photos of photo album **1364** to be transmitted to television set-top box **104** from user device **102** or remote storage and causing a slideshow of the photos to commence).

[0133] In yet another example, transcribed user speech **1472** includes a reference to the "last" photo ("Display the last photo on the kitchen television."). The particular photo referenced can be unclear from the speech input alone. The content shown on user device **102** can again be used to disambiguate the user request. In particular, the content shown in interface **1360** of FIG. **13** can be used to determine the user intent from the command to display the "last" photo. The listing of photos and videos in interface **1360** includes two individual photos **1366**. The appearance of photos **1366** in interface **1360**—and particularly the order of appearance of photos **1366** within the interface—can be used to determine the user's intent from utter the "last" photo. In particular, the user's reference to the "last" photo can be resolved to photo **1366** appearing at the bottom of interface **1360** (dated Jun. 21, 2014). In response to user speech **1472**, therefore, the virtual assistant can cause the last photo **1366** shown in interface **1360** to be displayed (e.g., by causing the last photo **1366** to be transmitted to television set-top box **104** from user device **102** or remote storage and causing the photo to be displayed).

[0134] In other examples, a user can reference media content shown in interface **1360** in a variety of other ways (e.g., the last couple of photos, all of the videos, all of the photos, the graduation album, the graduation video, the photo from June 21st, etc.), and user intent can similarly be determined based on displayed content. It should be appreciated that metadata associated with displayed content (e.g., timestamps, location information, titles, descriptions, etc.), fuzzy

matching techniques, synonym matching, and the like can further be used in conjunction with displayed content to determine user intent. Displayed content and data associated with it can thus be used to disambiguate user requests and determine user intent.

[0135] It should be understood that any type of displayed content in any application interface of any application can be used in determining user intent. For example, images displayed on a webpage in an Internet browser application can be referenced in speech input, and the displayed webpage content can be analyzed to identify the desired images. Similarly, a music track in a list of music in a music application can be referenced in speech input by title, genre, artist, band name, or the like, and the displayed content in the music application (and associated metadata in some examples) can be used to determine user intent from the speech input. As discussed above, the determined user intent can then be used to cause media display or playback via another device, such as via television set-top box **104**.

[0136] In some examples, user identification, user authentication, and/or device authentication can be employed to determine whether media control can be permitted, determine media content available for display, determine access permissions, and the like. For example, it can be determined whether a particular user device (e.g., user device **102**) is authorized to control media on, for example, television set-top box **104**. A user device can be authorized based on a registration, pairing, trust determination, passcode, security question, system setup, or the like. In response to determining that a particular user device is authorized, attempts to control television set-top box **104** can be permitted (e.g., media content can be played in response to determining that a requesting device is authorized to control media). In contrast, media control commands or requests from unauthorized devices can be ignored, and/or users of such devices can be prompted to register their devices for use in controlling a particular television set-top box **104**.

[0137] In another example, a particular user can be identified, and personal data associated with the user can be used to determine user intent of requests. For example, a user can be identified based on speech input, such as by voice recognition using a voiceprint of the user. In some examples, users can utter a particular phrase that is analyzed for voice recognition. In other examples, speech input requests directed to the virtual assistant can be analyzed using voice recognition to identify the speaker. A user can also be identified based on the source of the speech input sample (e.g., on a user's personal device **102**). A user can also be identified based on passwords, passcodes, menu selection, or the like. Speech input received from the user can then be interpreted based on personal data of the identified user. For example, user intent of speech input can be determined based on previous requests from the user, media content owned by the user, media content stored on the user's device, user preferences, user settings, user demographics (e.g., languages spoken, etc.), user profile information, user payment methods, or a variety of other personal information associated with a particular identified user. For instance, speech input referencing a favorites list or the like can be disambiguated based on personal data, and the user's personal favorites list can be identified. Speech input referencing "my" photos, "my" videos, "my" shows, or the like can likewise be disambiguated based on user identification to correctly identify photos, videos, and shows associated with the identified user (e.g., photos stored on a personal user device or the like). Similarly, speech input requesting purchase of content can be disambiguated to determine that the identified user's payment method should be charged for the purchase (as opposed to another user's payment method).

[0138] In some examples, user authentication can be used to determine whether a user is allowed to access media content, purchase media content, or the like. For example, voice recognition can be used to verify the identity of a particular user (e.g., using their voiceprint) to permit the user to make purchases using the user's payment method. Similarly, passwords or the like can be used to authenticate the user to permit purchases. In another example, voice recognition can be used to verify the identity of a particular user to determine whether the user is allowed to watch a particular program (e.g., a program having a particular parental guideline rating, a movie having a particular age suitability rating, or the like). For instance, a child's request for a particular program can be denied based on voice recognition indicating that the requester is not an authorized user able to view such content (e.g., a parent). In other examples, voice recognition can be used to determine whether users have access to particular subscription content (e.g., restricting access to premium channel content based on voice recognition). In some examples, users can utter a particular phrase that is analyzed for voice recognition. In other examples, speech input requests directed to the virtual assistant can be analyzed using voice recognition to identify the speaker. Certain media content can thus be played in response to first determining that a user is authorized in any of a variety of ways.

[0139] FIG. **15** illustrates exemplary virtual assistant interactions with results on a mobile user device and a media display device. In some examples, a virtual assistant can provide information and control on more than one device, such as on user device **102** as well as on television set-top box **104**. Moreover, in some examples, the same virtual assistant interface used for control and information on user device **102** can be used to issue requests for controlling media on television set-top box **104**. As such, the virtual assistant system can determine whether to display results or execute tasks on user device **102** or on television set-top box **104**. In some examples, when employing user device **102** to control television set-top box **104**, virtual assistant interface intrusion on a display associated with television set-top box **104** (e.g., display **112**) can be minimized by displaying information on user device **102** (e.g., on touchscreen **246**). In other examples, virtual assistant information can be displayed on display **112** alone, or virtual assistant information can be displayed on both user device **102** and display **112**.

[0140] In some examples, a determination can be made as to whether results of a virtual assistant query should be displayed on user device **102** directly or on display **112** associated with television set-top box **104**. In one example, in response to determining that the user intent of a query includes a request for information, an informational response can be displayed on user device **102**. In another example, in response to determining that the user intent of a query includes a request to play media content, media content responsive to the query can be played via television set-top box **104**.

[0141] FIG. **15** illustrates virtual assistant interface **1254** with a conversational dialog example between a virtual assistant and a user. Assistant greeting **1256** can prompt the user to make a request. In the first query, transcribed user speech **1574** (which can also be typed or entered in other ways)

includes a request for an informational answer associated with displayed media content. In particular, transcribed user speech **1574** inquires who is playing in a soccer game that may be, for example, shown on an interface on user device **102** (e.g., listed in interface **1150** of FIG. **11**) or on display **112** (e.g., listed in interface **510** of FIG. **5** or playing as video **726** on display **112** of FIG. **7**B). The user intent of transcribed user speech **1574** can be determined based on displayed media content. For example, the particular soccer game in question can be identified based on content shown on user device **102** or on display **112**. The user intent of transcribed user speech **1574** can include obtaining an informational answer detailing the teams playing in the soccer game identified based on the displayed content. In response to determining that the user intent includes a request for an informational answer, the system can determine to display the response within interface **1254** in FIG. **15** (as opposed to on display **112**). The response to the query can, in some examples, be determined based on metadata associated with the displayed content (e.g., based on a description of the soccer game in a television listing). As shown, assistant response **1576** can thus be displayed on touchscreen **246** of user device **102** in interface **1254**, identifying teams Alpha and Zeta as playing in the game. Accordingly, in some examples, an informational response can be displayed within interface **1254** on user device **102** based on determining that a query includes an informational request.

[0142] The second query in interface **1254**, however, includes a media request. In particular, transcribed user speech **1578** requests changing displayed media content to "the game." The user intent of transcribed user speech **1578** can be determined based on displayed content (e.g., to identify which game the user desires), such as a game listed in interface **510** of FIG. **5**, a game listed in interface **1150** of FIG. **11**, a game referenced in previous queries (e.g., in transcribed user speech **1574**), or the like. The user intent of transcribed user speech **1578** can thus include changing displayed content to a particular game—here, the soccer game with teams Alpha and Zeta. In one example, the game can be displayed on user device **102**. In other examples, however, based on the query including a request to play media content, the game can be shown via television set-top box **104**. In particular, in response to determining that the user intent includes a request to play media content, the system can determine to display the media content result via television set-top box **104** on display **112** (as opposed to within interface **1254** in FIG. **15**). In some examples, a response or paraphrase confirming the virtual assistant's intended action can be shown in interface **1254** or on display **112** (e.g., "Changing to the soccer game.").

[0143] FIG. **16** illustrates exemplary virtual assistant interactions with media results on a media display device and a mobile user device. In some examples, a virtual assistant can provide access to media on both user device **102** and television set-top box **104**. Moreover, in some examples, the same virtual assistant interface used for media on user device **102** can be used to issue requests for media on television set-top box **104**. As such, the virtual assistant system can determine whether to display media results on user device **102** or on display **112** via television set-top box **104**.

[0144] In some examples, a determination can be made as to whether to display media on device **102** or on display **112** based on media result format, user preference, default settings, an express command in the request itself, or the like. For example, the format of a media result to a query can be used to determine on which device to display the media result by default (e.g., without specific instructions). A television program can be better suited for display on a television, a large format video can be better suited for display on a television, thumbnail photos can be better suited for display on a user device, small format web videos can be better suited for display on a user device, and various other media formats can be better suited for display on either a relatively large television screen or a relatively small user device display. Thus, in response to a determination that media content should be displayed on a particular display (e.g., based on media format), the media content can be displayed on that particular display by default.

[0145] FIG. **16** illustrates virtual assistant interface **1254** with examples of queries related to playing or displaying media content. Assistant greeting **1256** can prompt the user to make a request. In the first query, transcribed user speech **1680** includes a request to show a soccer game. As in the examples discussed above, the user intent of transcribed user speech **1680** can be determined based on displayed content (e.g., to identify which game the user desires), such as a game listed in interface **510** of FIG. **5**, a game listed in interface **1150** of FIG. **11**, a game referenced in previous queries, or the like. The user intent of transcribed user speech **1680** can thus include displaying a particular soccer game that may, for example, be aired on television. In response to determining that the user intent includes a request to display media that is formatted for television (e.g., a televised soccer game), the system can automatically determine to display the desired media on display **112** via television set-top box **104** (as opposed to on user device **102** itself). The virtual assistant system can then cause television set-top box **104** to tune to the soccer game and show it on display **112** (e.g., by executing the necessary tasks and/or sending the appropriate commands).

[0146] In the second query, however, transcribed user speech **1682** includes a request to show pictures of players of a team (e.g., pictures of "Team Alpha"). As in the examples discussed above, the user intent of transcribed user speech **1682** can be determined. The user intent of transcribed user speech **1682** can include performing a search (e.g., a web search) for pictures associated with "Team Alpha," and displaying the resulting pictures. In response to determining that the user intent includes a request to display media that may be presented in thumbnail format, or media associated with a web search, or other non-specific media without a particular format, the system can automatically determine to display the desired media result on touchscreen **246** in interface **1254** of user device **102** (as opposed to displaying the resulting pictures on display **112** via television set-top box **104**). For example, as shown, thumbnail photos **1684** can be displayed within interface **1254** on user device **102** in response to the user's query. The virtual assistant system can thus cause media of a certain format, or media that might be presented in a certain format (e.g., in a group of thumbnails), to be displayed on user device **102** by default.

[0147] It should be appreciated that, in some examples, the soccer game referenced in user speech **1680** can be shown on user device **102**, and photos **1684** can be shown on display **112** via television set-top box **104**. The default device for display, however, can be determined automatically based on media format, thereby simplifying media commands for the user. In other examples, the default device for displaying requested media content can be determined based on user preferences, default settings, the device used most recently to

display content, voice recognition to identify a user and a device associated with that user, or the like. For example, a user can set a preference or a default configuration can be set to display certain types of content (e.g., videos, slideshows, television programs, etc.) on display **112** via television set-top box **104** and other types of content (e.g., thumbnails, photos, web videos, etc.) on touchscreen **246** of user device **102**. Similarly, preferences or default configurations can be set to respond to certain queries by displaying content on one device or the other. In another example, all content can be displayed on user device **102** unless the user instructs otherwise.

[0148] In still other examples, a user query can include a command to display content on a particular display. For example, user speech **1472** of FIG. **14** includes a command to display a photo on the kitchen television. As a result, the system can cause display of the photo on a television display associated with the user's kitchen as opposed to displaying the photo on user device **102**. In other examples, a user can dictate which display device to use in a variety of other ways (e.g., on TV, on the big screen, in the living room, in the bedroom, on my tablet, on my phone, etc.). The display device to use for displaying media content results of virtual assistant queries can thus be determined in a variety of different ways.

[0149] FIG. **17** illustrates exemplary media device control based on proximity. In some examples, users may have multiple televisions and television set-top boxes within the same household or on the same network. For example, a household may have a television and set-top box set in the living room, another set in the bedroom, and another set in the kitchen. In other examples, multiple set-top boxes can be connected to the same network, such as a common network in an apartment or office building. Although users can pair, connect, or otherwise authorize remote control **106** and user device **102** for a particular set-top box to avoid unauthorized access, in other examples, remote controls and/or user devices can be used to control more than one set-top box. A user can, for example, use a single user device **102** to control a set-top box in the bedroom, in the living room, and in the kitchen. A user can also, for example, use a single user device **102** to control their own set-top box in their own apartment, as well as control a neighbor's set-top box in a neighbor's apartment (e.g., sharing content from user device **102** with the neighbor, such as showing a slideshow on the neighbor's TV of photos stored on user device **102**). Because the user can use a single user device **102** to control multiple different set-top boxes, the system can determine to which set-top box of multiple set-top boxes to send commands. Likewise, because a household can have multiple remote controls **106** that can operate multiple set-top boxes, the system can similarly determine to which set-top box of multiple set-top boxes to send commands.

[0150] In one example, proximity of devices can be used to determine to which of multiple set-top boxes to send commands (or on which display to show requested media content). A proximity can be determined between a user device **102** or remote control **106** and each of multiple set-top boxes. Issued commands can then be sent to the nearest set-top box (or requested media content can be displayed on the nearest display). Proximity can be determined (or at least approximated) in any of a variety of ways, such as time-of-flight measurements (e.g., using radio frequency), Bluetooth LE, electronic ping signals, proximity sensors, sound travel measurements, or the like. Measured or approximated distances can then be compared, and the device with the shortest distance can be issued the command (e.g., the nearest set-top box).

[0151] FIG. **17** illustrates multi-device system **1790** including first set-top box **1792** with first display **1786** and second set-top box **1794** with second display **1788**. In one example, a user can issue a command from user device **102** to display media content (e.g., without necessarily specifying where or on which device). Distance **1795** to first set-top box **1792** and distance **1796** to second set-top box **1794** can then be determined (or approximated). As shown, distance **1796** can be greater than distance **1795**. Based on proximity, the command from user device **102** can be issued to first set-top box **1792** as the nearest device and the likeliest to match the user's intent. In some examples, a single remote control **106** can also be used to control more than one set-top box. The desired device for control at a given time can be determined based on proximity. Distance **1797** to second set-top box **1794** and distance **1798** to first set-top box **1792** can be determined (or approximated). As shown, distance **1798** can be greater than distance **1797**. Based on proximity, commands from remote control **106** can be issued to second set-top box **1794** as the nearest device and the likeliest to match the user's intent. Distance measurements can be refreshed regularly or with each command to accommodate, for example, a user moving to a different room and desiring to control a different device.

[0152] It should be understood that a user can specify a different device for a command, in some cases overriding proximity. For example, a list of available display devices can be displayed on user device **102** (e.g., listing first display **1786** and second display **1788** by setup name, designated room, or the like, or listing first set-top box **1792** and second set-top box **1794** by setup name, designated room, or the like). A user can select one of the devices from the list, and commands can then be sent to the selected device. Requests for media content issued at user device **102** can then be handled by displaying the desired media on the selected device. In other examples, users can speak the desired device as part of a spoken command (e.g., show the game on the kitchen television, change to the cartoon channel in the living room, etc.).

[0153] In still other examples, the default device for showing requested media content can be determined based on status information associated with a particular device. For example, it can be determined whether headphones (or a headset) are attached to user device **102**. In response to determining that headphones are attached to user device **102** when a request to display media content is received, the requested content can be displayed on user device **102** by default (e.g., assuming the user is consuming content on user device **102** and not on a television). In response to determining that headphones are not attached to user device **102** when a request to display media content is received, the requested content can be displayed on either user device **102** or on a television according to any of the various determination methods discussed herein. Other device status information can similarly be used to determine whether requested media content should be displayed on user device **102** or a set-top box **104**, such as ambient lighting around user device **102** or set-top box **104**, proximity of other devices to user device **102** or set-top box **104**, orientation of user device **102** (e.g., landscape orientation can be more likely to indicate desired viewing on user device **102**), display status of set-top box **104** (e.g., in a sleep mode), time since the last interaction on a

particular device, or any of a variety of other status indicators for user device **102** and/or set-top box **104**.

[0154] FIG. **18** illustrates exemplary process **1800** for controlling television interactions using a virtual assistant and multiple user devices. At block **1802**, speech input can be received from a user at a first device with a first display. For example, speech input can be received from a user at user device **102** or remote control **106** of system **100**. The first display can include touchscreen **246** of user device **102** or a display associated with remote control **106** in some examples.

[0155] At block **1804**, user intent can be determined from the speech input based on content displayed on the first display. For example, content such as television programs **1152** in interface **1150** of FIG. **11** or photos and videos in interface **1360** of FIG. **13** can be analyzed and used to determine user intent for speech input. In some examples, a user can refer to content shown on the first display in ambiguous ways, and the references can be disambiguated by analyzing the content shown on the first display to resolve the references (e.g., determining the user intent for "that" video, "that" album, "that" game, or the like), as discussed above with reference to FIG. **12** and FIG. **14**.

[0156] Referring again to process **1800** of FIG. **18**, at block **1806**, media content can be determined based on the user intent. For example, a particular video, photo, photo album, television program, sporting event, music track, or the like can be identified based on the user intent. In the example of FIG. **11** and FIG. **12** discussed above, for instance, the particular soccer game shown on channel five can be identified based on the user intent referring to "that" soccer game shown in interface **1150** of FIG. **11**. In the examples of FIG. **13** and FIG. **14** discussed above, the particular video **1362** titled "Graduation Video," the particular photo album **1364** titled "Graduation Album," or a particular photo **1366** can be identified based on the user intent determined from the speech input examples of FIG. **14**.

[0157] Referring again to process **1800** of FIG. **18**, at block **1808**, the media content can be played on a second device associated with a second display. For example, the determined media content can be played via television set-top box **104** on display **112** with speakers **111**. Playing the media content can include tuning to a particular television channel, playing a particular video, showing a slideshow of photos, displaying a particular photo, playing a particular audio track, or the like on television set-top box **104** or another device.

[0158] In some examples, a determination can be made as to whether responses to speech input directed to a virtual assistant should be displayed on a first display associated with a first device (e.g., user device **102**) or a second display associated with a second device (e.g., television set-top box **104**). For example, as discussed above with reference to FIG. **15** and FIG. **16**, informational answers or media content suited for display on a smaller screen can be displayed on user device **102**, while media responses or media content suited for display on a larger screen can be displayed on a display associated with set-top box **104**. As discussed above with reference to FIG. **17**, in some examples, the distance between user device **102** and multiple set-top boxes can be used to determine on which set-top box to play media content or to which set-top box to issue commands. Various other determinations can similarly be made to provide a convenient and user-friendly experience where multiple devices may be interacting.

[0159] In some examples, as content shown on user device **102** can be used to inform interpretations of speech input as discussed above, content shown on display **112** can likewise be used to inform interpretations of speech input. In particular, content shown on a display associated with television set-top box **104** can be used along with metadata associated with that content to determine user intent from speech input, disambiguate user queries, respond to content-related queries, or the like.

[0160] FIG. **19** illustrates exemplary speech input interface **484** (described above) with a virtual assistant query about video **480** shown in the background. In some examples, user queries can include questions about media content shown on display **112**. For example, transcription **1916** includes a query requesting identification of actresses ("Who are those actresses?"). Content shown on display **112**—along with metadata or other descriptive information about the content—can be used to determine user intent from speech input relating to that content as well as to determine responses to queries (responses including both informational responses as well as media responses providing media selections to the user). For example, video **480**, a description of video **480**, a character and actor list for video **480**, rating information for video **480**, genre information for video **480**, and a variety of other descriptive information associated with video **480** can be used to disambiguate user requests and determine responses to user queries. Associated metadata can include, for example, identifying information for character **1910**, character **1912**, and character **1914** (e.g., character names along with the names of the actresses who play the characters). Metadata for any other content can similarly include a title, a description, a list of characters, a list of actors, a list of players, a genre, producer names, director names, or a display schedule associated with the content shown on the display or the viewing history of media content on the display (e.g., recently displayed media).

[0161] In one example, a user query directed to a virtual assistant can include an ambiguous reference to something shown on display **112**. Transcription **1916**, for example, includes a reference to "those" actresses ("Who are those actresses?"). The particular actresses the user is asking about can be unclear from the speech input alone. In some examples, however, the content shown on display **112** and associated metadata can be used to disambiguate user requests and determine user intent. In the illustrated example, the content shown on display **112** can be used to determine the user intent from the reference to "those" actresses. In one example, television set-top box **104** can identify playing content along with details associated with the content. In this instance, television set-top box **104** can identify the title of video **480** along with a variety of descriptive content. In other examples, a television show, sporting event, or other content can be shown that can be used in conjunction with associated metadata to determine user intent. In addition, in any of the various examples discussed herein, speech recognition results and intent determination can weight terms associated with displayed content higher than alternatives. For example, actor names for on-screen characters can be weighted higher while those actors appear on screen (or while a show is playing in which they appear), which can provide for accurate speech recognition and intent determination of likely user requests associated with displayed content.

[0162] In one example, a character and/or actor list associated with video **480** can be used to identify all or the most

prominent actresses appearing in video **480**, which might include actresses **1910**, **1912**, and **1914**. The identified actresses can be returned as a possible result (including fewer or additional actresses if the metadata resolution is coarse). In another example, however, metadata associated with video **480** can include an identification of which actors and actresses appear on screen at a given time, and the actresses appearing at the time of the query can be determined from that metadata (e.g., specifically identifying actresses **1910**, **1912**, and **1914**). In yet another example, a facial recognition application can be used to identify actresses **1910**, **1912**, and **1914** from the images shown on display **112**. In still other examples, various other metadata associated with video **480** and various other recognition approaches can be used to identify the user's likely intent in referring to "those" actresses.

[0163] In some examples, the content shown on display **112** can change during submission of a query and determination of a response. As such, a viewing history of media content can be used to determine user intent and determine the response to a query. For example, should video **480** move to another view (e.g., with other characters) before a response to the query is generated, the result of the query can be determined based on the user's view at the time the query was spoken (e.g., the characters shown on screen at the time the user initiated the query). In some instances, a user might pause playing media to issue a query, and the content shown when paused can be used with associated metadata to determine user intent and a response to the query.

[0164] Given the determined user intent, a result of the query can be provided to the user. FIG. **20** illustrates exemplary assistant response interface **2018** including assistant response **2020**, which can include the response determined from the query of transcription **1916** of FIG. **19**. Assistant response **2020** can include, as shown, a listing of each actress's name and her associated character in video **480** ("Actress Jennifer Jones plays the character Blanche; actress Elizabeth Arnold plays the character Julia; and actress Whitney Davidson plays the character Melissa."). The listed actresses and characters in response **2020** can correspond to characters **1910**, **1912**, and **1914** appearing on display **112**. As noted above, in some examples, the content shown on display **112** can change during submission of a query and determination of a response. As such, response **2020** can include information about content or characters that may no longer appear on display **112**.

[0165] As with other interfaces displayed on display **112**, assistant response interface **2018** can occupy a minimal amount of screen real estate while providing sufficient space to convey the desired information. In some examples, as with other text displayed in interfaces on display **112**, assistant response **2020** can be scrolled up into the position shown in FIG. **20** from the bottom of display **112**, displayed for a certain amount of time (e.g., a delay based on the length of the response), and scrolled up out of view. In other examples, interface **2018** can be slid downward out of view after a delay.

[0166] FIG. **21** and FIG. **22** illustrate another example of determining user intent and responding to a query based on content shown on display **112**. FIG. **21** illustrates exemplary speech input interface **484** with a virtual assistant query for media content associated with video **480**. In some examples, user queries can include a request for media content associated with media shown on display **112**. For example, a user can request other movies, television programs, sporting events, or the like associated with particular media based, for example, on a character, actor, genre, or the like. For example, transcription **2122** includes a query requesting other media associated with an actress in video **480**, referenced by her character's name in video **480** ("What else is Blanche in?"). Content shown on display **112**—along with metadata or other descriptive information about the content—can again be used to determine user intent from speech input relating to that content as well as to determine responses to queries (either informational or resulting in media selections).

[0167] In some examples, a user query directed to a virtual assistant can include an ambiguous reference using the name of a character, the name of an actor, the name of a program, the name of player, or the like. Without the context of the content shown on display **112** and its associated metadata, such references may be difficult to resolve accurately. Transcription **2122**, for example, includes a reference to a character named "Blanche" from video **480**. The particular actress or other individual the user is asking about can be unclear from the speech input alone. In some examples, however, the content shown on display **112** and associated metadata can be used to disambiguate user requests and determine user intent. In the illustrated example, the content shown on display **112** and associated metadata can be used to determine the user intent from the character name "Blanche." In this instance, a character list associated with video **480** can be used to determine that "Blanche" likely refers to the character "Blanche" in video **480**. In another example, detailed metadata and/or facial recognition can be used to determine that a character with the name "Blanche" appears on the screen (or appeared on the screen at the initiation of the user's query), making the actress associated with that character the likeliest intention of the user's query. For example, it can be determined that characters **1910**, **1912**, and **1914** appear on display **112** (or appeared on display **112** at the initiation of the user's query), and their associated character names can then be referenced to determine the user intent of the query referencing the character Blanche. An actor list can then be used to identify the actress who plays Blanche, and a search can be conducted to identify other media in which the identified actress appears.

[0168] Given the determined user intent (e.g., resolution of the character reference "Blanche") and the determination of the result of the query (e.g., other media associated with the actress who plays "Blanche"), a response can be provided to the user. FIG. **22** illustrates exemplary assistant response interface **2224** including assistant text response **2226** and selectable video links **2228**, which can be responsive to the query of transcription **2122** of FIG. **21**. Assistant text response **2226** can include, as shown, a paraphrase of the user request introducing selectable video links **2228**. Assistant text response **2226** can also include an indication of the disambiguation of the user's query—in particular, identifying actress Jennifer Jones as playing the character Blanche in video **480**. Such a paraphrase can confirm to the user that the virtual assistant correctly interpreted the user's query and is providing the desired result.

[0169] Assistant response interface **2224** can also include selectable video links **2228**. In some examples, various types of media content can be provided as results to a virtual assistant query, including movies (e.g., Movie A and Movie B of interface **2224**). Media content displayed as a result of a query can include media that may be available to the user for consumption (for free, for purchase, or as part of a subscription). A user can select displayed media to view or consume the

resulting content. For instance, a user can select one of selectable video links **2228** (e.g., using a remote control, voice command, or the like) to watch one of the other movies in which actress Jennifer Jones appears. In response to selection of one of selectable video links **2228**, the video associated with the selection can be played, replacing video **480** on display **112**. Thus, displayed media content and associated metadata can be used to determine user intent from speech input, and, in some examples, playable media can be provided as a result.

[0170] It should be understood that a user can reference actors, players, characters, locations, teams, sporting event details, movie subjects, or a variety of other information associated with displayed content in forming queries, and the virtual assistant system can similarly disambiguate such requests and determine user intent based on displayed content and associated metadata. Likewise, it should be understood that, in some examples, results can include media suggestions associated with the query, such as a movie, television show, or sporting event associated with a person who is the subject of a query (whether or not the user specifically requests such media content).

[0171] Moreover, in some examples, user queries can include requests for information associated with media content itself, such as queries about a character, an episode, a movie plot, a previous scene, or the like. As with the examples discussed above, displayed content and associated metadata can be used to determine user intent from such queries and determine a response. For instance, a user might request a description of a character (e.g., "What does Blanche do in this movie?"). The virtual assistant system can then identify from metadata associated with displayed content the requested information about the character, such as a character description or role (e.g., "Blanche is one of a group of lawyers and is known as a troublemaker in Hartford."). Similarly, a user might request an episode synopsis (e.g., "What happened in the last episode?"), and the virtual assistant system can search for and provide a description of the episode.

[0172] In some examples, content displayed on display **112** can include menu content, and such menu content can similarly be used to determine user intent of speech input and responses to user queries. FIGS. **23**A-**23**B illustrate exemplary pages of a program menu **830**. FIG. **23**A illustrates a first page of media options **832**, and FIG. **23**B illustrates a second page of media options **832** (which can include a consecutive next page in a listing of content that extends beyond a single page).

[0173] In one example, a user request to play content can include an ambiguous reference to something shown on display **112** in menu **830**. For example, a user viewing menu **830** can request to watch "that" soccer game, "that" basketball game, the vacuum advertisement, the law show, or the like. The particular program desired can be unclear from the speech input alone. In some examples, however, the content shown on display **112** can be used to disambiguate user requests and determine user intent. In the illustrated example, the media options in menu **830** (along with metadata associated with the media options in some examples) can be used to determine the user intent from commands including ambiguous references. For example, "that" soccer game can be resolved to the soccer game on the sports channel. "That" basketball game can be resolved to the basketball game on the college sports channel. The vacuum advertisement can be resolved to the paid programming show (e.g., based on metadata associated with the show describing a vacuum). The law show can be resolved to the courtroom drama based on metadata associated with the show and/or synonym matching, fuzzy matching, or other matching techniques. The appearance of the various media options **832** in menu **830** on display **112** can thus be used to disambiguate user requests.

[0174] In some examples, displayed menus can be navigated with a cursor, joystick, arrows, buttons, gestures, or the like. In such instances, a focus can be shown for a selected item. For example, a selected item can be shown in bold, underlined, outlined with a border, in larger size than other menu items, with a shadow, with a reflection, with a glow, and/or with any other features to emphasize which menu item is selected and has focus. For example, selected media option **2330** in FIG. **23**A can have focus as the currently selected media option, and is shown with large, underlined type and a border.

[0175] In some examples, a request to play content or select a menu item can include an ambiguous reference to a menu item that has focus. For example, a user viewing menu **830** of FIG. **23**A can request to play "that" show (e.g., "Play that show."). Similarly, a user could request various other commands associated with a menu item having focus, such as play, delete, hide, remind me to watch that, record that, or the like. The particular menu item or show that is desired can be unclear from the speech input alone. The content shown on display **112**, however, can be used to disambiguate user requests and determine user intent. In particular, the fact that selected media option **2330** has focus in menu **830** can be used to identify the desired media subject of any of the commands referring to "that" show, commands without subjects (e.g., play, delete, hide, etc.), or any other ambiguous commands referring to the media content having focus. A menu item having focus can thus be used in determining user intent from speech input.

[0176] As with a viewing history of media content that can be used to disambiguate a user request (e.g., content displayed at the time a user initiated a request but since having passed), previously displayed menu or search result content can similarly be used to disambiguate later user requests after moving on, for example, to later menu or search result content. For example, FIG. **23**B illustrates a second page of menu **830** with additional media options **832**. A user can advance to the second page illustrated in FIG. **23**B but refer back to content shown in the first page illustrated in FIG. **23**A (e.g., media options **832** shown in FIG. **23**A). For example, despite having moved on to the second page of menu **830**, a user can request to watch "that" soccer game, "that" basketball game, or the law show—all of which are media options **832** recently displayed on a previous page of menu **830**. Such references can be ambiguous, but the recently displayed menu content from the first page of menu **830** can be used to determine the user intent. In particular, the recently displayed media options **832** of FIG. **23**A can be analyzed to identify the specific soccer game, basketball game, or courtroom drama referred to in the ambiguous example requests. In some examples, results can be biased based on how recently content was displayed (e.g., weighting the most recently viewed page of results over results viewed earlier). In this manner, the viewing history of what was recently shown on display **112** can be used to determine user intent. It should be understood that any recently displayed content can be used, such as previously displayed search results, previously displayed programs, previously displayed menus, or the like. This can allow users to

refer back to something they saw earlier without having to find and navigate to the specific view in which they saw it.

[0177] In still other examples, various display cues shown in a menu or results list on display **112** can be used to disambiguate user requests and determine user intent. FIG. **24** illustrates an exemplary media menu divided into categories, one of which has focus (movies). FIG. **24** illustrates category interface **2440**, which can include a carousel-style interface of categorized media options including TV options **2442**, movie options **2444**, and music options **2446**. As shown, the music category is only partially displayed, and the carousel interface can be shifted to display additional content to the right (e.g., as indicated by the arrow) as though rotating the media in a carousel. In the illustrated example, the movies category has focus as indicated by the underlined title and border, although focus can be indicated in any of a variety of other ways (e.g., making the category larger to appear closer to the user than other categories, adding a glow, etc.).

[0178] In some examples, a request to play content or select a menu item can include an ambiguous reference to a menu item in a group of items (such as a category). For example, a user viewing category interface **2440** can request to play the soccer show ("Play the soccer show."). The particular menu item or show that is desired can be unclear from the speech input alone. Moreover, the query can resolve to more than one show that is displayed on display **112**. For example, the request for the soccer show might refer to either the soccer game listed in the TV programs category or the soccer movie listed in the movies category. The content shown on display **112**—including display cues—can be used to disambiguate user requests and determine user intent. In particular, the fact that the movies category has focus in category interface **2440** can be used to identify the particular soccer show that is desired, which is likely the soccer movie given the focus on the movies category. A category of media (or any other grouping of media) having focus as shown on display **112** can thus be used in determining user intent from speech input. It should also be appreciated that users can make various other requests associated with categories, such as requesting display of certain categorical content (e.g., show me comedy movies, show me honor movies, etc.).

[0179] In other examples, a user can refer to menu or media items shown on display **112** in a variety of other ways, and user intent can similarly be determined based on displayed content. It should be appreciated that metadata associated with displayed content (e.g., TV program descriptions, movie descriptions, etc.), fuzzy matching techniques, synonym matching, and the like can further be used in conjunction with displayed content to determine user intent from speech input. User requests in a variety of forms—including natural language requests—can thus be accommodated and user intent can be determined according to the various examples discussed herein.

[0180] It should be understood that content displayed on display **112** can be used alone or in conjunction with content displayed on user device **102** or on a display associated with remote control **106** in determining user intent. Likewise, it should be understood that virtual assistant queries can be received at any of a variety of devices communicatively coupled to television set-top box **104**, and content displayed on display **112** can be used to determine user intent regardless of which device receives the query. Results of queries can likewise be displayed on display **112** or on another display (e.g., on user device **102**).

[0181] In addition, in any of the various examples discussed herein, the virtual assistant system can navigate menus and select menu options without requiring a user to specifically open menus and navigate to menu items. For example, a menu of options might appear after selecting media content or a menu button, such as selecting a movie option **2444** in FIG. **24**. Menu options might include playing the media as well as alternatives to simply playing the media, such as setting a reminder to watch the media later, setting up a recording of the media, adding media to a favorites list, hiding media from further view, or the like. While a user is viewing content above a menu or content that has a sub-menu option, the user can issue virtual assistant commands that would otherwise require navigating to the menu or sub-menu to select. For example, a user viewing category interface **2440** of FIG. **24** can issue any menu command associated with a movie option **2444** without opening the associated menu manually. For instance, the user might request to add the soccer movie to a favorites list, record the nightly news, and set up a reminder to watch Movie B without ever navigating to the menus or sub-menus associated with those media options where such commands might be available. The virtual assistant system can thus navigate menus and sub-menus in order to execute commands on behalf of the user, whether or not those menu options appear on display **112**. This can simplify user requests and reduce the number of clicks or selections a user must make to achieve desired menu functionality.

[0182] FIG. **25** illustrates exemplary process **2500** for controlling television interactions using media content shown on a display and a viewing history of media content. At block **2502**, speech input can be received from a user, the speech input including a query associated with content shown on a television display. For example, the speech input can include a query about a character, actor, movie, television program, sporting event, player, or the like appearing on display **112** of system **100** (shown by television set-top box **104**). Transcription **1916** of FIG. **19**, for example, includes a query associated with actresses shown in video **480** on display **112**. Similarly, transcription **2122** of FIG. **21** includes a query associated with a character in video **480** shown on display **112**. The speech input can also include a query associated with menu or search content appearing on display **112**, such as a query to select a particular menu item or get information about a particular search result. For example, displayed menu content can include media options **832** of menu **830** in FIG. **23**A and FIG. **23**B. Displayed menu content can likewise include TV options **2442**, movie options **2444**, and/or music options **2446** appearing in category interface **2440** of FIG. **24**.

[0183] Referring again to process **2500** of FIG. **25**, at block **2504**, user intent of the query can be determined based on the content shown and a viewing history of media content. For example, user intent can be determined based on a displayed or recently displayed scene of a television program, sporting event, movie, or the like. User intent can also be determined based on displayed or recently displayed menu or search content. Displayed content can also be analyzed along with metadata associated with the content to determine user intent. For example, the content shown and described with reference to FIGS. **19**, **21**, **23**A, **23**B, and **24** can be used alone or in conjunction with metadata associated with the displayed content to determine user intent.

[0184] At block **2506**, a result of the query can be displayed based on the determined user intent. For example, a result similar to assistant response **2020** in assistant response inter-

face **2018** of FIG. **20** can be displayed on display **112**. In another example, text and selectable media can be provided as a result, such as assistant text response **2226** and selectable video links **2228** in assistant response interface **2224** shown in FIG. **22**. In yet another example, displaying the result of the query can include displaying or playing selected media content (e.g., playing a selected video on display **112** via television set-top box **104**). User intent can thus be determined from speech input in a variety of ways using displayed content and associated metadata as context.

[0185] In some examples, virtual assistant query suggestions can be provided to a user to, for example, inform the user of available queries, suggest content that the user may enjoy, teach the user how to use the system, encourage the user to find additional media content for consumption, or the like. In some examples, query suggestions can include generic suggestions of possible commands (e.g., find comedies, show me the TV guide, search for action movies, turn on closed captioning, etc.). In other examples, query suggestions can include targeted suggestions related to displayed content (e.g., add this show to a watch list, share this show via social media, show me the soundtrack of this movie, show me the book that this guest is selling, show me the trailer for the movie that guest is plugging, etc.), user preferences (e.g., closed captioning use, etc.), user-owned content, content stored on a user's device, notifications, alerts, a viewing history of media content (e.g., recently displayed menu items, recently displayed scenes of a show, recent actor appearances, etc.), or the like. Suggestions can be displayed on any device, including on display **112** via television set-top box **104**, on user device **102**, or on a display associated with remote control **106**. In addition, suggestions can be determined based on which devices are nearby and/or in communication with television set-top box **104** at a particular time (e.g., suggesting content from devices of the users in the room watching TV at a particular time). In other examples, suggestions can be determined based on a variety of other contextual information, including the time of day, crowd-sourced information (e.g., popular shows being watched at a given time), shows that are live (e.g., live sporting events), a viewing history of media content (e.g., the last several shows that were watched, a recently viewed set of search results, a recently viewed group of media options, etc.), or any of a variety of other contextual information.

[0186] FIG. **26** illustrates exemplary suggestions interface **2650** including content-based virtual assistant query suggestions **2652**. In one example, query suggestions can be provided in an interface such as interface **2650** in response to input received from a user requesting suggestions. Input requesting query suggestions can be received, for example, from user device **102** or remote control **106**. In some examples, the input can include a button press, a double click of a button, a menu selection, a voice command (e.g., show me some suggestions, what can you do for me, what are some options, etc.), or the like received at user device **102** or remote control **106**. For instance, a user can double click a physical button on remote control **106** to request query suggestions, or can double click a physical or virtual button on user device **102** when viewing an interface associated with television set-top box **104** to request query suggestions.

[0187] Suggestions interface **2650** can be displayed over a moving image, such as video **480**, or over any other background content (e.g., a menu, a still image, a paused video, etc.). As with other interfaces discussed herein, suggestions interface **2650** can be animated to slide up from the bottom of display **112**, and can occupy a minimal amount of space while sufficiently conveying the desired information so as to limit interference with video **480** in the background. In other examples, a larger interface of suggestions can be provided when the background content is still (e.g., a paused video, a menu, an image, etc.).

[0188] In some examples, virtual assistant query suggestions can be determined based on displayed media content or a viewing history of media content (e.g., a movie, television show, sporting event, recently viewed show, recently viewed menu, recently viewed scene of a movie, recent scene of a playing television episode, etc.). For example, FIG. **26** illustrates content-based suggestions **2652**, which can be determined based on displayed video **480** shown in the background with characters **1910**, **1912**, and **1914** appearing on display **112**. Metadata associated with displayed content (e.g., descriptive details of the media content) can also be used to determine query suggestions. Metadata can include a variety of information associated with displayed content, including a show title, a character list, an actor list, an episode description, a team roster, a team ranking, a show synopsis, movie details, plot descriptions, director names, producer names, times of actor appearance, sports standings, sports scores, genre, season episode listing, related media content, or a variety of other associated information. For example, metadata associated with video **480** can include the character names of characters **1910**, **1912**, and **1914** along with the actresses who play those characters. Metadata can also include a description of the plot of video **480**, a description of a previous or next episode (where video **480** is a television episode in a series), or the like.

[0189] FIG. **26** illustrates a variety of content-based suggestions **2652** that can be shown in suggestions interface **2650** based on video **480** and metadata associated with video **480**. For example, character **1910** of video **480** can be named "Blanche," and the character name can be used to formulate a query suggestion for information about the character Blanche or the actress who plays that character (e.g., "Who is the actress that plays Blanche?"). Character **1910** can be identified from metadata associated with video **480** (e.g., a character list, an actor list, times associated with actor appearances, etc.). In other examples, facial recognition can be used to identify actresses and/or characters appearing on display **112** at a given time. Various other query suggestions can be provided associated with a character in the media itself, such as queries relating to a character's role, profile, relationship to other characters, or the like.

[0190] In another example, an actor or actress appearing on display **112** can be identified (e.g., based on metadata and/or facial recognition), and query suggestions associated with that actor or actress can be provided. Such query suggestions can include role(s) played, acting awards, age, other media in which they appear, history, family members, relationships, or any of a variety of other details about an actor or actress. For example, character **1914** can be played by an actress named Whitney Davidson, and the actress's name Whitney Davidson can be used to formulate a query suggestion to identify other movies, television programs, or other media in which the actress Whitney Davidson appears (e.g., "What else is Whitney Davidson in?").

[0191] In other examples, details about a show can be used to formulate query suggestions. An episode synopsis, plot summary, episode list, episode titles, series titles, or the like

can be used to formulate query suggestions. For example, a suggestion can be provided to describe what happened in the last episode of a television program (e.g., "What happened in the last episode?"), to which the virtual assistant system can provide as a response an episode synopsis from the prior episode identified based on the episode currently shown on display **112** (and its associated metadata). In another example, a suggestion can be provided to set up a recording for the next episode, which can be accomplished by the system identifying the next episode based on the currently playing episode shown on display **112**. In yet another example, a suggestion can be provided to get information about the current episode or show appearing on display **112**, and the title of the show obtained from metadata can be used to formulate the query suggestion (e.g., "What is this episode of 'Their Show' about?" or "What is 'Their Show' about?").

[0192] In another example, category, genre, rating, awards, descriptions, or the like associated with displayed content can be used to formulate query suggestions. For example, video **480** can correspond to a television program described as a comedy having female lead characters. A query suggestion can be formulated from this information to identify other shows with similar characteristics (e.g., "Find me other comedies with female leads."). In other examples, suggestions can be determined based on user subscriptions, content available for playback (e.g., content on television set-top box **104**, content on user device **102**, content available for streaming, etc.), or the like. For example, potential query suggestions can be filtered based on whether informational or media results are available. Query suggestions that might not result in playable media content or informational answers can be excluded, and/or query suggestions with readily available informational answers or playable media content can be provided (or weighted more heavily in determining which suggestions to provide). Displayed content and associated metadata can thus be used in a variety of ways to determine query suggestions.

[0193] FIG. **27** illustrates exemplary selection interface **2754** for confirming selection of a suggested query. In some examples, users can select displayed query suggestions by speaking the queries, selecting them with a button, navigating to them with a cursor, or the like. In response to a selection, the selected suggestion can be briefly displayed in a confirming interface, such as selection interface **2754**. In one example, selected suggestion **2756** can be animated to move from wherever it appeared in suggestions interface **2650** to the position shown in FIG. **27** next to command receipt confirmation **490** (e.g., as shown by the arrow), and other unselected suggestions can be hidden from the display.

[0194] FIGS. **28**A-**28**B illustrate exemplary virtual assistant answer interface **2862** based on a selected query. In some examples, informational answers to a selected query can be displayed in an answer interface, such as answer interface **2862**. In switching from either suggestions interface **2650** or selection interface **2754**, transition interface **2858** can be shown as illustrated in FIG. **28**A. In particular, previously displayed content within the interface can be scrolled upward out of the interface as the next content scrolls upward from the bottom of display **112**. Selected suggestion **2756**, for example, can be slid or scrolled upward until it disappears at the top edge of the virtual assistant interface, and assistant result **2860** can be slid or scrolled upward from the bottom of display **112** until it arrives at the position shown in FIG. **28**B.

[0195] Answer interface **2862** can include informational answers and/or media results responsive to a selected query suggestion (or responsive to any other query). For example, in response to selected query suggestion **2756**, assistant result **2860** can be determined and provided. In particular, in response to a request for a synopsis of a prior episode, the prior episode can be identified based on displayed content, and an associated description or synopsis can be identified and provided to the user. In the illustrated example, assistant result **2860** can describe a previous episode of the program corresponding to video **480** on display **112** (e.g., "In episode **203** of 'Their Show,' Blanche gets invited to a college psychology class as a guest speaker. Julia and Melissa show up unannounced and cause a stir."). Informational answers and media results (e.g., selectable video links) can also be presented in any of the other ways discussed herein, or results can be presented in various other ways (e.g., speaking answers aloud, playing content immediately, showing an animation, displaying an image, etc.).

[0196] In another example, a notification or alert can be used to determine virtual assistant query suggestions. FIG. **29** illustrates a media content notification **2964** (although any notification can be taken into account in determining suggestions) and suggestions interface **2650** with both notification-based suggestions **2966** and content-based suggestions **2652** (which can include some of the same concepts as discussed above with reference to FIG. **26**). In some examples, the content of a notification can be analyzed to identify relevant media related names, titles, subjects, actions, or the like. In the illustrated example, notification **2964** includes an alert notifying the user about alternative media content available for display—specifically that a sporting event is live, and the content of the game may be of interest to the user (e.g., "Team Zeta and Team Alpha are tied with five minutes remaining in the game."). In some examples, notifications can be displayed momentarily at the top of display **112**. Notifications can be slid down from the top of display **112** (as indicated by the arrow) into the position shown in FIG. **29**, displayed for a certain amount of time, and slid back up to disappear again at the top of display **112**.

[0197] Notifications or alerts can notify the user of a variety of information, such as available alternative media content (e.g., alternatives to what may be shown currently on display **112**), available live television programs, newly downloaded media content, recently added subscription content, suggestions received from friends, receipt of media sent from another device, or the like. Notifications can also be personalized based on a household or an identified user watching media (e.g., identified based on user authentication using account selections, voice recognition, passwords, etc.). In one example, the system can interrupt a show and display a notification based on likely desired content, such as displaying notification **2964** for a user who—based on a user profile, favorite team(s), preferred sport(s), viewing history, and the like—can be likely to desire the content of the notification. For example, sporting event scores, game status, time remaining, and the like can be obtained from a sport data feed, news outlet, social media discussions, or the like, and can be used to identify possible alternative media content for notifying the user.

[0198] In other examples, popular media content (e.g., across many users) can be provided via alerts or notifications to suggest alternatives to currently viewed content (e.g., notifying a user that a popular show or a show in a genre the user likes just started or is otherwise available for viewing). In the illustrated example, the user might follow one or both of

Team Zeta and Team Alpha (or might follow soccer or a particular sport, league, etc.). The system can determine that available live content matches the user's preferences (e.g., a game on another channel matches a user's preferences, the game has little time remaining, and the score is close). The system can then determine to alert the user via notification **2964** of the likely desired content. In some examples, a user can select notification **2964** (or a link within notification **2964**) to switch to the suggested content (e.g., using a remote control button, cursor, spoken request, etc.).

[0199] Virtual assistant query suggestions can be determined based on notifications by analyzing notification content to identify relevant media related terms, names, titles, subjects, actions, or the like. The identified information can then be used to formulate appropriate virtual assistant query suggestions, such as notification-based suggestions **2966** based on notification **2964**. For example, a notification about an exciting end of a live sporting event can be displayed. Should the user then request query suggestions, suggestions interface **2650** can be displayed, including query suggestions to view the sporting event, inquire about team statistics, or find content related to the notification (e.g., change to the Zeta/Alpha game, what are team Zeta's stats, what other soccer games are on, etc.). Based on the particular terms of interest identified in the notification, various other query suggestions can likewise be determined and provided to the user.

[0200] Virtual assistant query suggestions related to media content (e.g., for consumption via television set-top box **104**) can also be determined from content on a user device, and suggestions can also be provided on a user device. In some examples, playable device content can be identified on user devices that are connected to or in communication with television set-top box **104**. FIG. **30** illustrates user device **102** with exemplary picture and video content in interface **1360**. A determination can be made as to what content is available for playback on a user device, or what content is likely to be desired for playback. For example, playable media **3068** can be identified based on an active application (e.g., a photos and videos application), or can be identified based on stored content whether displayed on interface **1360** or not (e.g., content can be identified from an active application in some examples or without being displayed at a given time in other examples). Playable media **3068** can include, for example, video **1362**, photo album **1364**, and photos **1366**, each of which can include personal user content that can be transmitted to television set-top box **104** for display or playback. In other examples, any photo, video, music, game interface, application interface, or other media content stored or displayed on user device **102** can be identified and used for determining query suggestions.

[0201] With playable media **3068** identified, virtual assistant query suggestions can be determined and provided to the user. FIG. **31** illustrates exemplary TV assistant interface **3170** on user device **102** with virtual assistant query suggestions based on playable user device content and based on video content shown on a separate display (e.g., display **112** associated with television set-top box **104**). TV assistant interface **3170** can include a virtual assistant interface specifically for interacting with media content and/or television set-top box **104**. Users can request query suggestions on user device **102** by, for example, a double click of a physical button when viewing interface **3170**. Other inputs can similarly be used to indicate a request for query suggestions. As shown, assistant greeting **3172** can introduce the provided query suggestions (e.g., "Here are some suggestions for controlling your TV experience.").

[0202] Virtual assistant query suggestions provided on user device **102** can include suggestions based on a variety of source devices as well as general suggestions. For example, device-based suggestions **3174** can include query suggestions based on content stored on user device **102** (including content displayed on user device **102**). Content-based suggestions **2652** can be based on content displayed on display **112** associated with television set-top box **104**. General suggestions **3176** can include general suggestions that may not be associated with particular media content or a particular device with media content.

[0203] Device-based suggestions **3174** can be determined, for example, based on playable content identified on user device **102** (e.g., videos, music, photographs, game interfaces, application interfaces, etc.). In the illustrated example, device-based suggestions **3174** can be determined based on playable media **3068** shown in FIG. **30**. For example, given that photo album **1364** was identified as playable media **3068**, the details of photo album **1364** can be used to formulate a query. The system can identify the content as an album of multiple photos that can be shown in a slideshow, and can then use the title of the album (in some instances) to formulate a query suggestion to show a slideshow of the particular album of photos (e.g., "Show a slideshow of 'Graduation Album' from your photos."). In some examples, the suggestion can include an indication of the source of the content (e.g., "from your photos," "from Jennifer's phone," "from Daniel's tablet," etc.). The suggestion can also use other details to refer to particular content, such as a suggestion to view a photograph from a particular date (e.g., display your photo from June 21st). In another example, video **1362** can be identified as playable media **3068**, and the title of the video (or other identifying information) can be used to formulate a query suggestion to play the video (e.g., "Show 'Graduation Video' from your videos.").

[0204] In other examples, content available on other connected devices can be identified and used to formulate virtual assistant query suggestions. For example, content from each of two user devices **102** connected to a common television set-top box **104** can be identified and used in formulating virtual assistant query suggestions. In some examples, users can select which content to make visible to the system for sharing, and can hide other content from the system so as not to include it in query suggestions or otherwise make it available for playback.

[0205] Content-based suggestions **2652** shown in interface **3170** of FIG. **31** can be determined, for example, based on content displayed on display **112** associated with television set-top box **104**. In some examples, content-based suggestions **2652** can be determined in the same manner as described above with reference to FIG. **26**. In the illustrated example, content-based suggestions **2652** shown in FIG. **31** can be based on video **480** shown on display **112** (e.g., as in FIG. **26**). In this manner, virtual assistant query suggestions can be derived based on content that is displayed or available on any number of connected devices. In addition to targeted suggestions, general suggestions **3176** can be predetermined and provided (e.g., show me the guide, what sports are on, what's on channel three, etc.).

[0206] FIG. **32** illustrates exemplary suggestions interface **2650** with connected device-based suggestions **3275** along

with content-based suggestions **2652** shown on display **112** associated with television set-top box **104**. In some examples, content-based suggestions **2652** can be determined in the same manner as described above with reference to FIG. **26**. As noted above, virtual assistant query suggestions can be formulated based on content on any number of connected devices, and the suggestions can be provided on any number of connected devices. FIG. **32** illustrates connected device-based suggestions **3275** that can be derived from content on user device **102**. For example, playable content can be identified on user device **102**, such as photo and video content shown in interface **1360** as playable media **3068** in FIG. **30**. The identified playable content on user device **102** can then be used to formulate suggestions that can be displayed on display **112** associated with television set-top box **104**. In some examples, connected device-based suggestions **3275** can be determined in the same manner as device-based suggestions **3174** described above with reference to FIG. **31**. In addition, as noted above, in some examples identifying source information can be included in a suggestion, such as "from Jake's phone" as shown in connected device-based suggestions **3275**. Virtual assistant query suggestions provided on one device can thus be derived based on content from another device (e.g., displayed content, stored content, etc.). It should be appreciated that a connected device can include a remote storage device accessible to television set-top box **104** and/or user device **102** (e.g., accessing media content stored in the cloud to formulate suggestions).

[0207] It should be understood that any combination of virtual assistant query suggestions from various sources can be provided in response to a request for suggestions. For example, suggestions from various sources can be combined randomly, or can be presented based on popularity, user preference, selection history, or the like. Moreover, queries can be determined in a variety of other ways and presented based on a variety of other factors, such as a query history, a user preference, a query popularity, or the like. In addition, in some examples, query suggestions can be cycled automatically by replacing displayed suggestions with new alternative suggestions after a delay. It should further be understood that users can select displayed suggestions on any interface by, for example, tapping on a touchscreen, speaking the query, selecting a query with navigation keys, selecting a query with a button, selecting a query with a cursor, or the like, and an associated response can then be provided (e.g., an informational and/or media response).

[0208] In any of the various examples, virtual assistant query suggestions can also be filtered based on available content. For example, potential query suggestions that would result in unavailable media content (e.g., no cable subscription) or that may not have an associated informational answer can be disqualified as suggestions and held back from being displayed. On the other hand, potential query suggestions that would result in immediately playable media content to which the user has access can be weighted over other potential suggestions or otherwise biased for display. In this manner, the availability of media content for user viewing can also be used in determining virtual assistant query suggestions for display.

[0209] In addition, in any of the various examples, pre-loaded query answers can be provided instead of or in addition to suggestions (e.g., in suggestions interface **2650**). Such pre-loaded query answers can be selected and provided based on personal use and/or current context. For example, a user watching a particular program can tap a button, double-click a button, long-press a button, or the like to receive suggestions. Instead of or in addition to query suggestions, context-based information can be provided automatically, such as identifying a playing song or soundtrack (e.g., "This song is Performance Piece"), identifying cast members of a currently playing episode (e.g., "Actress Janet Quinn plays Genevieve"), identifying similar media (e.g., "Show Q is similar to this"), or providing results of any of the other queries discussed herein.

[0210] Moreover, affordances can be provided in any of the various interfaces for users to rate media content to inform the virtual assistant of user preferences (e.g., a selectable rating scale). In other examples, users can speak rating information as a natural language command (e.g., "I love this," "I hate this," "I don't like this show," etc.). In still other examples, in any of the various interfaces illustrated and described herein, a variety of other functional and informational elements can be provided. For example, interfaces can further include links to important functions and places, such as search links, purchase links, media links, and the like. In another example, interfaces can further include recommendations of what else to watch next based on currently playing content (e.g., selecting similar content). In yet another example, interfaces can further include recommendations of what else to watch next based on personalized taste and/or recent activity (e.g., selecting content based on user ratings, user-entered preferences, recently watched programs, etc.). In still other examples, interfaces can further include instructions for user interactions (e.g., "Press and hold to talk to the Virtual Assistant," "Tap once to get suggestions," etc.). In some examples, providing pre-loaded answers, suggestions, or the like can provide an enjoyable user experience while also making content readily available to a wide variety of users (e.g., to users of various skill levels irrespective of language or other control barriers).

[0211] FIG. **33** illustrates exemplary process **3300** for suggesting virtual assistant interactions for controlling media content (e.g., virtual assistant queries). At block **3302**, media content can be displayed on a display. For example, as shown in FIG. **26**, video **480** can be displayed on display **112** via television set-top box **104**, or interface **1360** can be displayed on touchscreen **246** of user device **102** as shown in FIG. **30**. At block **3304**, an input can be received from a user. The input can include a request for virtual assistant query suggestions. The input can include a button press, a double click of a button, a menu selection, a spoken query for suggestions, or the like.

[0212] At block **3306**, virtual assistant queries can be determined based on the media content and/or a viewing history of media content. For example, virtual assistant queries can be determined based on a displayed program, menu, application, list of media content, notification, or the like. In one example, content-based suggestions **2652** can be determined based on video **480** and associated metadata as described with reference to FIG. **26**. In another example, notification-based suggestions **2966** can be determined based on notification **2964** as described with reference to FIG. **29**. In yet another example, device-based suggestions **3174** can be determined based on playable media **3068** on user device **102** as described with reference to FIG. **30** and FIG. **31**. In still other examples, connected device-based suggestions **3275** can be determined based on playable media **3068** on user device **102** as described with reference to FIG. **32**.

[0213] Referring again to process **3300** of FIG. **33**, at block **3308**, the virtual assistant queries can be displayed on the display. For example, determined query suggestions can be displayed as shown in and described with reference to FIGS. **26**, **27**, **29**, **31**, and **32**. As discussed above, query suggestions can be determined and displayed based on a variety of other information. Moreover, virtual assistant query suggestions provided on one display can be derived based on content from another device with another display. Targeted virtual assistant query suggestions can thus be provided to users, thereby assisting users to learn of potential queries as well as providing desirable content suggestions, among other benefits.

[0214] In addition, in any of the various examples discussed herein, various aspects can be personalized for a particular user. User data, including contacts, preferences, location, favorite media, and the like, can be used to interpret voice commands and facilitate user interaction with the various devices discussed herein. The various processes discussed herein can also be modified in various other ways according to user preferences, contacts, text, usage history, profile data, demographics, or the like. In addition, such preferences and settings can be updated over time based on user interactions (e.g., frequently uttered commands, frequently selected applications, etc.). Gathering and use of user data that is available from various sources can be used to improve the delivery to users of invitational content or any other content that may be of interest to them. The present disclosure contemplates that in some instances, this gathered data can include personal information data that uniquely identifies or can be used to contact or locate a specific person. Such personal information data can include demographic data, location-based data, telephone numbers, email addresses, home addresses, or any other identifying information.

[0215] The present disclosure recognizes that the use of such personal information data, in the present technology, can be used to the benefit of users. For example, the personal information data can be used to deliver targeted content that is of greater interest to the user. Accordingly, use of such personal information data enables calculated control of the delivered content. Further, other uses for personal information data that benefit the user are also contemplated by the present disclosure.

[0216] The present disclosure further contemplates that the entities responsible for the collection, analysis, disclosure, transfer, storage, or other use of such personal information data will comply with well-established privacy policies and/or privacy practices. In particular, such entities should implement and consistently use privacy policies and practices that are generally recognized as meeting or exceeding industry or governmental requirements for maintaining personal information data as private and secure. For example, personal information from users should be collected for legitimate and reasonable uses of the entity and not shared or sold outside of those legitimate uses. Further, such collection should occur only after receiving the informed consent of the users. Additionally, such entities would take any needed steps for safeguarding and securing access to such personal information data and ensuring that others with access to the personal information data adhere to their privacy policies and procedures. Further, such entities can subject themselves to evaluation by third parties to certify their adherence to widely accepted privacy policies and practices.

[0217] Despite the foregoing, the present disclosure also contemplates examples in which users selectively block the use of, or access to, personal information data. That is, the present disclosure contemplates that hardware and/or software elements can be provided to prevent or block access to such personal information data. For example, in the case of advertisement delivery services, the present technology can be configured to allow users to select to "opt in" or "opt out" of participation in the collection of personal information data during registration for services. In another example, users can select not to provide location information for targeted content delivery services. In yet another example, users can select not to provide precise location information, but permit the transfer of location zone information.

[0218] Therefore, although the present disclosure broadly covers use of personal information data to implement one or more various disclosed examples, the present disclosure also contemplates that the various examples can also be implemented without the need for accessing such personal information data. That is, the various examples of the present technology are not rendered inoperable due to the lack of all or a portion of such personal information data. For example, content can be selected and delivered to users by inferring preferences based on non-personal information data or a bare minimum amount of personal information, such as the content being requested by the device associated with a user, other non-personal information available to the content delivery services, or publicly available information.

[0219] In accordance with some examples, FIG. **34** shows a functional block diagram of an electronic device **3400** configured in accordance with the principles of various described examples to, for example, control television interactions using a virtual assistant and display associated information using different interfaces. The functional blocks of the device can be implemented by hardware, software, or a combination of hardware and software to carry out the principles of the various described examples. It is understood by persons of skill in the art that the functional blocks described in FIG. **34** can be combined or separated into sub-blocks to implement the principles of the various described examples. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein.

[0220] As shown in FIG. **34**, electronic device **3400** can include a display unit **3402** configured to display media, interfaces, and other content (e.g., display **112**, touchscreen **246**, or the like). Electronic device **3400** can further include input unit **3404** configured to receive information, such as speech input, tactile input, gesture input, and the like (e.g., a microphone, a receiver, a touchscreen, a button, or the like). Electronic device **3400** can further include processing unit **3406** coupled to display unit **3402** and input unit **3404**. In some examples, processing unit **3406** can include a speech input receiving unit **3408**, a media content determining unit **3410**, a first user interface displaying unit **3412**, a selection receiving unit **3414**, and a second user interface displaying unit **3416**.

[0221] Processing unit **3406** can be configured to receive speech input from a user (e.g., via input unit **3404**). Processing unit **3406** can be further configured to determine (e.g., using media content determining unit **3410**) media content based on the speech input. Processing unit **3406** can be further configured to display (e.g., on display unit **3402** using first user interface displaying unit **3412**) a first user interface having a first size, wherein the first user interface comprises one or more selectable links to the media content. Processing unit

3406 can be further configured to receive (e.g., from input unit **3404** using selection receiving unit **3414**) a selection of one of the one or more selectable links. Processing unit **3406** can be further configured to, in response to the selection, display (e.g., on display unit **3402** using second user interface displaying unit **3416**) a second user interface having a second size larger than the first size, wherein the second user interface comprises the media content associated with the selection.

[0222] In some examples, the first user interface (e.g., of first user interface displaying unit **3412**) expands into the second user interface (e.g., of second user interface displaying unit **3416**) in response to the selection (e.g., of selection receiving unit **3414**). In other examples, the first user interface is overlaid on playing media content. In one example, the second user interface is overlaid on playing media content. In another example, the speech input (e.g., of speech input receiving unit **3408** from input unit **3404**) comprises a query, and the media content (e.g., of media content determining unit **3410**) comprises a result of the query. In still another example, the first user interface comprises a link to results of the query beyond the one or more selectable links to the media content. In other examples, the query comprises a query about weather, and the first user interface comprises a link to media content associated with the query about the weather. In another example, the query comprises a location, and the link to the media content associated with the query about the weather comprises a link to a portion of media content associated with weather at the location.

[0223] In some examples, in response to the selection, processing unit **3406** can be configured to play the media content associated with the selection. In one example, the media content comprises a movie. In another example, the media content comprises a television show. In another example, the media content comprises a sporting event. In some examples, the second user interface (e.g., of second user interface displaying unit **3416**) comprises a description of the media content associated with the selection. In other examples, the first user interface comprises a link to purchase media content.

[0224] Processing unit **3406** can be further configured to receive additional speech input from the user (e.g., via input unit **3404**), wherein the additional speech input comprises a query associated with displayed content. Processing unit **3406** can be further configured to determine a response to the query associated with the displayed content based on metadata associated with the displayed content. Processing unit **3406** can be further configured to, in response to receiving the additional speech input, display (e.g., on display unit **3402**) a third user interface, wherein the third user interface comprises the determined response to the query associated with the displayed content.

[0225] Processing unit **3406** can be further configured to receive an indication to initiate receipt of speech input (e.g., via input unit **3404**). Processing unit **3406** can be further configured to, in response to receiving the indication, display a readiness confirmation (e.g., on display unit **3402**). Processing unit **3406** can be further configured to, in response to receiving the speech input, display a listening confirmation. Processing unit **3406** can be further configured to detect the end of the speech input, and, in response to detecting the end of the speech input, display a processing confirmation. In some examples, processing unit **3406** can be further configured to display a transcription of the speech input.

[0226] In some examples, electronic device **3400** comprises a television. In other examples, electronic device **3400** comprises a television set-top box. In other examples, electronic device **3400** comprises a remote control. In still other examples, electronic device **3400** comprises a mobile telephone.

[0227] In one example, the one or more selectable links in the first user interface (e.g., of first user interface displaying unit **3412**) comprise moving images associated with the media content. In some examples, the moving images associated with the media content comprise live feeds of the media content. In other examples, the one or more selectable links in the first user interface comprise still images associated with the media content.

[0228] In some examples, processing unit **3406** can be further configured to determine whether currently displayed content comprises a moving image or a control menu; in response to a determination that currently displayed content comprises a moving image, select a small size as the first size for the first user interface (e.g., of first user interface displaying unit **3412**); and, in response to a determination that currently displayed content comprises a control menu, select a large size, larger than the small size, as the first size for the first user interface (e.g., of first user interface displaying unit **3412**). In other examples, processing unit **3406** can be further configured to determine alternative media content for display based on one or more of a user preference, a show popularity, and a status of a live sporting event, and to display a notification comprising the determined alternative media content.

[0229] In accordance with some examples, FIG. **35** shows a functional block diagram of an electronic device **3500** configured in accordance with the principles of various described examples to, for example, control television interactions using a virtual assistant and multiple user devices. The functional blocks of the device can be implemented by hardware, software, or a combination of hardware and software to carry out the principles of the various described examples. It is understood by persons of skill in the art that the functional blocks described in FIG. **35** can be combined or separated into sub-blocks to implement the principles of the various described examples. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein.

[0230] As shown in FIG. **35**, electronic device **3500** can include a display unit **3502** configured to display media, interfaces, and other content (e.g., display **112**, touchscreen **246**, or the like). Electronic device **3500** can further include input unit **3504** configured to receive information, such as speech input, tactile input, gesture input, and the like (e.g., a microphone, a receiver, a touchscreen, a button, or the like). Electronic device **3500** can further include processing unit **3506** coupled to display unit **3502** and input unit **3504**. In some examples, processing unit **3506** can include a speech input receiving unit **3508**, a user intent determining unit **3510**, a media content determining unit **3512**, and a media content playing unit **3514**.

[0231] Processing unit **3506** can be configured to receive (e.g., from input unit **3504** using speech input receiving unit **3508**) speech input from a user at a first device (e.g., device **3500**) having a first display (e.g., display unit **3502** in some examples). Processing unit **3506** can be further configured to determine (e.g., using user intent determining unit **3510**) a user intent of the speech input based on content displayed on the first display. Processing unit **3506** can be further config-

ured to determine (e.g., using media content determining unit **3512**) media content based on the user intent. Processing unit **3506** can be further configured to play (e.g., using media content playing unit **3514**) the media content on a second device associated with a second display (e.g., display unit **3502** in some examples).

[0232] In one example, the first device comprises a remote control. In another example, the first device comprises a mobile telephone. In another example, the first device comprises a tablet computer. In some examples, the second device comprises a television set-top box. In other examples, the second display comprises a television.

[0233] In some examples, the content displayed on the first display comprises an application interface. In one example, the speech input (e.g., of speech input receiving unit **3508** from input unit **3504**) comprises a request to display media associated with the application interface. In one example, the media content comprises the media associated with the application interface. In another example, the application interface comprises a photo album, and the media comprises one or more photos in the photo album. In yet another example, the application interface comprises a list of one or more videos, and the media comprises one of the one or more videos. In still other examples, the application interface comprises a television program listing, and the media comprises a television program in the television program listing.

[0234] In some examples, processing unit **3506** can be further configured to determine whether the first device is authorized; wherein the media content is played on the second device in response to a determination that the first device is authorized. Processing unit **3506** can be further configured to identify the user based on the speech input, and determine (e.g., using user intent determining unit **3510**) the user intent of the speech input based on data associated with the identified user. Processing unit **3506** can be further configured to determine whether the user is authorized based on the speech input; wherein the media content is played on the second device in response to a determination that the user is an authorized user. In one example, determining whether the user is authorized comprises analyzing the speech input using voice recognition.

[0235] In other examples, processing unit **3506** can be further configured to, in response to determining that the user intent comprises a request for information, display information associated with the media content on the first display of the first device. Processing unit **3506** can be further configured to, in response to determining that the user intent comprises a request to play the media content, play the media content on the second device.

[0236] In some examples, the speech input comprises a request to play content on the second device, and the media content is played on the second device in response to the request to play content on the second device. Processing unit **3506** can be further configured to determine whether the determined media content should be displayed on the first display or the second display based on a media format, a user preference, or a default setting. In some examples, the media content is displayed on the second display in response to a determination that the determined media content should be displayed on the second display. In other examples, the media content is displayed on the first display in response to a determination that the determined media content should be displayed on the first display.

[0237] In other examples, processing unit **3506** can be further configured to determine a proximity of each of two or more devices, including the second device and a third device. In some examples, the media content is played on the second device associated with the second display based on the proximity of the second device relative to the proximity of the third device. In some examples, determining the proximity of each of the two or more devices comprises determining the proximity based on Bluetooth LE.

[0238] In some examples, processing unit **3506** can be further configured to display a list of display devices, including the second device associated with the second display, and receive a selection of the second device in the list of display devices. In one example, the media content is displayed on the second display in response to receiving the selection of the second device. Processing unit **3506** can be further configured to determine whether headphones are attached to the first device. Processing unit **3506** can be further configured to, in response to a determination that headphones are attached to the first device, display the media content on the first display. Processing unit **3506** can be further configured to, in response to a determination that headphones are not attached to the first device, display the media content on the second display. In other examples, processing unit **3506** can be further configured to determine alternative media content for display based on one or more of a user preference, a show popularity, and a status of a live sporting event, and to display a notification comprising the determined alternative media content.

[0239] In accordance with some examples, FIG. **36** shows a functional block diagram of an electronic device **3600** configured in accordance with the principles of various described examples to, for example, control television interactions using media content shown on a display and a viewing history of media content. The functional blocks of the device can be implemented by hardware, software, or a combination of hardware and software to carry out the principles of the various described examples. It is understood by persons of skill in the art that the functional blocks described in FIG. **36** can be combined or separated into sub-blocks to implement the principles of the various described examples. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein.

[0240] As shown in FIG. **36**, electronic device **3600** can include a display unit **3602** configured to display media, interfaces, and other content (e.g., display **112**, touchscreen **246**, or the like). Electronic device **3600** can further include input unit **3604** configured to receive information, such as speech input, tactile input, gesture input, and the like (e.g., a microphone, a receiver, a touchscreen, a button, or the like). Electronic device **3600** can further include processing unit **3606** coupled to display unit **3602** and input unit **3604**. In some examples, processing unit **3606** can include a speech input receiving unit **3608**, a user intent determining unit **3610**, and a query result displaying unit **3612**.

[0241] Processing unit **3606** can be configured to receive (e.g., from input unit **3604** using speech input receiving unit **3608**) speech input from a user, wherein the speech input comprises a query associated with content shown on a television display (e.g., display unit **3602** in some examples). Processing unit **3606** can be further configured to determine (e.g., using user intent determining unit **3610**) a user intent of the query based on one or more of the content shown on the television display and a viewing history of media content.

Processing unit **3606** can be further configured to display (e.g., using query result displaying unit **3612**) a result of the query based on the determined user intent.

[0242] In one example, the speech input is received at a remote control. In another example, the speech input is received at a mobile telephone. In some examples, the result of the query is displayed on the television display. In another example, the content shown on the television display comprises a movie. In yet another example, the content shown on the television display comprises a television show. In still another example, the content shown on the television display comprises a sporting event.

[0243] In some examples, the query comprises a request for information about a person associated with the content shown on the television display, and the result (e.g., of query result displaying unit **3612**) of the query comprises information about the person. In one example, the result of the query comprises media content associated with the person. In another example, the media content comprises one or more of a movie, a television show, or a sporting event associated with the person. In some examples, the query comprises a request for information about a character in the content shown on the television display, and the result of the query comprises information about the character or information about the actor who plays the character. In one example, the result of the query comprises media content associated with the actor who plays the character. In another example, the media content comprises one or more of a movie, a television show, or a sporting event associated with the actor who plays the character.

[0244] In some examples, processing unit **3606** can be further configured to determine the result of the query based on metadata associated with the content shown on the television display or the viewing history of media content. In one example, the metadata comprises one or more of a title, a description, a list of characters, a list of actors, a list of players, a genre, or a display schedule associated with the content shown on the television display or the viewing history of media content. In another example, the content shown on the television display comprises a list of media content, and the query comprises a request to display one of the items in the list. In yet another example, the content shown on the television display further comprises an item in the list of media content having focus, and determining (e.g., using user intent determining unit **3610**) the user intent of the query comprises identifying the item having focus. In some examples, processing unit **3606** can be further configured to determine (e.g., using user intent determining unit **3610**) the user intent of the query based on menu or search content recently displayed on the television display. In one example, the content shown on the television display comprises a page of listed media, and the recently displayed menu or search content comprises a previous page of listed media. In another example, the content shown on the television display comprises one or more categories of media, and one of the one or more categories of media has focus. In one example, processing unit **3606** can be further configured to determine (e.g., using user intent determining unit **3610**) the user intent of the query based on the one of the one or more categories of media having focus. In another example, the categories of media comprise movies, television programs, and music. In other examples, processing unit **3606** can be further configured to determine alternative media content for display based on one or more of a user preference, a show popularity, and a status of a live sporting event, and to display a notification comprising the determined alternative media content.

[0245] In accordance with some examples, FIG. **37** shows a functional block diagram of an electronic device **3700** configured in accordance with the principles of various described examples to, for example, suggest virtual assistant interactions for controlling media content. The functional blocks of the device can be implemented by hardware, software, or a combination of hardware and software to carry out the principles of the various described examples. It is understood by persons of skill in the art that the functional blocks described in FIG. **37** can be combined or separated into sub-blocks to implement the principles of the various described examples. Therefore, the description herein optionally supports any possible combination or separation or further definition of the functional blocks described herein.

[0246] As shown in FIG. **37**, electronic device **3700** can include a display unit **3702** configured to display media, interfaces, and other content (e.g., display **112**, touchscreen **246**, or the like). Electronic device **3700** can further include input unit **3704** configured to receive information, such as speech input, tactile input, gesture input, and the like (e.g., a microphone, a receiver, a touchscreen, a button, or the like). Electronic device **3700** can further include processing unit **3706** coupled to display unit **3702** and input unit **3704**. In some examples, processing unit **3706** can include a media content displaying unit **3708**, an input receiving unit **3710**, a query determining unit **3712**, and a query displaying unit **3714**.

[0247] Processing unit **3706** can be configured to display (e.g., using media content displaying unit **3708**) media content on a display (e.g., display unit **3702**). Processing unit **3706** can be further configured to receive (e.g., from input unit **3704** using input receiving unit **3710**) an input from a user. Processing unit **3706** can be further configured to determine (e.g., using query determining unit **3712**) one or more virtual assistant queries based on one or more of the media content and a viewing history of media content. Processing unit **3706** can be further configured to display (e.g., using query displaying unit **3714**) the one or more virtual assistant queries on the display.

[0248] In one example, the input is received from the user on a remote control. In another example, the input is received from the user on a mobile telephone. In some examples, the one or more virtual assistant queries are overlaid on a moving image. In another example, the input comprises a double click of a button. In one example, the media content comprises a movie. In another example, the media content comprises a television show. In yet another example, the media content comprises a sporting event.

[0249] In some examples, the one or more virtual assistant queries comprise a query about a person appearing in the media content. In other examples, the one or more virtual assistant queries comprise a query about a character appearing in the media content. In another example, the one or more virtual assistant queries comprise a query for media content associated with a person appearing in the media content. In some examples, the media content or the viewing history of media content comprise an episode of a television show, and the one or more virtual assistant queries comprise a query about another episode of the television show. In another example, the media content or the viewing history of media content comprise an episode of a television show, and the one or more virtual assistant queries comprise a request to set a

reminder to watch or record a subsequent episode of the media content. In still another example, the one or more virtual assistant queries comprise a query for descriptive details of the media content. In one example, the descriptive details comprise one or more of a show title, a character list, an actor list, an episode description, a team roster, a team ranking, or a show synopsis.

[0250] In some examples, processing unit **3706** can be further configured to receive a selection of one of the one or more virtual assistant queries. Processing unit **3706** can be further configured to display a result of the selected one of the one or more virtual assistant queries. In one example, determining the one or more virtual assistant queries comprises determining the one or more virtual assistant queries based on one or more of a query history, a user preference, or a query popularity. In another example, determining the one or more virtual assistant queries comprises determining the one or more virtual assistant queries based on media content available to the user for viewing. In yet another example, determining the one or more virtual assistant queries comprises determining the one or more virtual assistant queries based on a received notification. In still another example, determining the one or more virtual assistant queries comprises determining the one or more virtual assistant queries based on an active application. In other examples, processing unit **3706** can be further configured to determine alternative media content for display based on one or more of a user preference, a show popularity, and a status of a live sporting event, and to display a notification comprising the determined alternative media content.

[0251] Although examples have been fully described with reference to the accompanying drawings, it is to be noted that various changes and modifications will become apparent to those skilled in the art (e.g., modifying any of the systems or processes discussed herein according to the concepts described in relation to any other system or process discussed herein). Such changes and modifications are to be understood as being included within the scope of the various examples as defined by the appended claims.

What is claimed is:

1. A method for controlling television interactions using a virtual assistant, the method comprising:

at an electronic device:

receiving speech input from a user;

determining media content based on the speech input;

displaying a first user interface having a first size, wherein the first user interface comprises one or more selectable links to the media content;

receiving a selection of one of the one or more selectable links; and

in response to the selection, displaying a second user interface having a second size larger than the first size, wherein the second user interface comprises the media content associated with the selection.

2. The method of claim **1**, wherein the first user interface expands into the second user interface in response to the selection.

3. The method of claim **1**, wherein the first user interface is overlaid on playing media content.

4. The method of claim **1**, wherein the second user interface is overlaid on playing media content.

5. The method of claim **1**, wherein the speech input comprises a query, and the media content comprises a result of the query.

6. The method of claim **5**, wherein the first user interface comprises a link to results of the query beyond the one or more selectable links to the media content.

7. The method of claim **1**, further comprising:

in response to the selection, playing the media content associated with the selection.

8. The method of claim **1**, wherein the media content comprises a sporting event.

9. The method of claim **1**, wherein the second user interface comprises a description of the media content associated with the selection.

10. The method of claim **1**, wherein the first user interface comprises a link to purchase media content.

11. The method of claim **1**, further comprising:

receiving additional speech input from the user, wherein the additional speech input comprises a query associated with displayed content;

determining a response to the query associated with the displayed content based on metadata associated with the displayed content; and

in response to receiving the additional speech input, displaying a third user interface, wherein the third user interface comprises the determined response to the query associated with the displayed content.

12. The method of claim **1**, further comprising:

receiving an indication to initiate receipt of speech input; and

in response to receiving the indication, displaying a readiness confirmation.

13. The method of claim **1**, further comprising:

in response to receiving the speech input, displaying a listening confirmation.

14. The method of claim **1**, further comprising:

displaying a transcription of the speech input.

15. The method of claim **1**, wherein the electronic device comprises a television.

16. The method of claim **1**, wherein the electronic device comprises a television set-top box.

17. The method of claim **1**, wherein the electronic device comprises a remote control.

18. The method of claim **1**, wherein the electronic device comprises a mobile telephone.

19. The method of claim **1**, wherein the one or more selectable links in the first user interface comprise moving images associated with the media content.

20. The method of claim **19**, wherein the moving images associated with the media content comprise live feeds of the media content.

21. The method of claim **1**, further comprising:

determining whether currently displayed content comprises a moving image or a control menu;

in response to a determination that currently displayed content comprises a moving image, selecting a small size as the first size for the first user interface; and

in response to a determination that currently displayed content comprises a control menu, selecting a large size, larger than the small size, as the first size for the first user interface.

22. The method of claim **1**, further comprising:

determining alternative media content for display based on one or more of a user preference, a show popularity, and a status of a live sporting event; and

displaying a notification comprising the determined alternative media content.

23. A non-transitory computer-readable storage medium comprising computer-executable instructions for:

receiving speech input from a user;

determining media content based on the speech input;

displaying a first user interface having a first size, wherein the first user interface comprises one or more selectable links to the media content;

receiving a selection of one of the one or more selectable links; and

responsive to the selection, for displaying a second user interface having a second size larger than the first size, wherein the second user interface comprises the media content associated with the selection.

24. The non-transitory computer-readable storage medium of claim **23**, wherein the first user interface expands into the second user interface in response to the selection.

25. The non-transitory computer-readable storage medium of claim **23**, wherein the first user interface is overlaid on playing media content.

26. The non-transitory computer-readable storage medium of claim **23**, wherein the second user interface is overlaid on playing media content.

27. The non-transitory computer-readable storage medium of claim **23**, wherein the speech input comprises a query, and the media content comprises a result of the query.

28. The non-transitory computer-readable storage medium of claim **27**, wherein the first user interface comprises a link to results of the query beyond the one or more selectable links to the media content.

29. A system for controlling television interactions using a virtual assistant, the system comprising:

one or more processors;

memory; and

one or more programs, wherein the one or more programs are stored in the memory and configured to be executed by the one or more processors, the one or more programs including instructions for:

receiving speech input from a user;

determining media content based on the speech input;

displaying a first user interface having a first size, wherein the first user interface comprises one or more selectable links to the media content;

receiving a selection of one of the one or more selectable links; and

responsive to the selection, for displaying a second user interface having a second size larger than the first size, wherein the second user interface comprises the media content associated with the selection.

30. The system of claim **29**, wherein the first user interface expands into the second user interface in response to the selection.

31. The system of claim **29**, wherein the first user interface is overlaid on playing media content.

32. The system of claim **29**, wherein the second user interface is overlaid on playing media content.

33. The system of claim **29**, wherein the speech input comprises a query, and the media content comprises a result of the query.

34. The system of claim **33**, wherein the first user interface comprises a link to results of the query beyond the one or more selectable links to the media content.

* * * * *